附件 2

广播电视和网络视听区块链技术应用白皮书（2020）

——内容审核篇

国家广播电视总局科技司

2020 年 10 月

前 言

区块链作为继大数据、云计算、人工智能、虚拟现实、5G 等技术后又一项对未来信息化发展产生重大影响的新兴技术，有望推动人类从信息互联网时代步入价值互联网时代，在全球科技创新和产业变革中的重要作用日趋突显。区块链技术应用已延伸到数字金融、物联网、智能制造、供应链管理、数字资产交易等多个领域。全球主要国家都在加快布局区块链技术发展，我国也将区块链提升到核心技术自主创新重要突破口的国家战略高度。中共中央政治局 2019 年 10 月 24 日就区块链技术发展现状和趋势进行第十八次集体学习。习近平总书记在主持学习时强调，区块链技术的集成应用在新的技术革新和产业变革中起着重要作用。我们要把区块链作为核心技术自主创新的重要突破口，明确主攻方向，加大投入力度，着力攻克一批关键核心技术，加快推动区块链技术和产业创新发展。随后，各地陆续颁布与区块链相关政策文件，“区块链+”成为众多行业和地区重点扶植领域。

2019 年 11 月 18 日，广电总局党组学习贯彻习近平总书记的讲话精神，提出积极推动区块链技术在广播电视和网络视听领域创新应用的要求，明确指出要密切跟踪和研究区块链技术发展现状和趋势，提高运用和管理区块链技术能力，使区块链在媒体融合和广播电视提质增效等方面发挥更大作用。目前，区块链技术在广播电视和网络视听领域的应用仍处于探索和起步阶段，相关单位在版权保护、内容审核、用户激励、可信数据共享等相关场景应用区块链技术进行了有益的尝试，但区块链在行业内大规模应用的技术思路、成功案例还有待开发。著名咨询公司高德纳（Gartner）发布的 2019 年区块链技术成熟度曲线也预测指出，区块链在媒体行业的成熟应用可能尚需 5～10 年。但另一方面，我国广播电视和网络视听行业具有服务主体多、节目内容量大、传播环节多、覆盖用户多等突出特点，利用区块链去中心化、互信任、防篡改、可追溯的特点来实现多方参与、资源共享、数据可信、成本节约、安全监督具有重要而迫切的现实意义和广阔的应用前景。有必要找准制约广播电视和网络视听创新发展的难点、痛点问题，结合区块链技术特点和支撑能力，分析适合我国广播电视和网络视听区块链技术应用的框架结构和应用领域，提出我国广播电视和网络视听区块链技术应用的推进思路，为行业转型升级注入新动能、激发新活力，加快推进广播电视和网络视听

高质量创新性发展。

为更好地推动区块链在广播电视和网络视听行业的创新应用，加强区块链技术在广播电视和网络视听行业应用的引导与规范，国家广播电视总局科技司组织广播电视科学研究院、广播电视规划院、中广电广播电影电视设计研究院等行业内外相关机构、科研单位、高等院校、企业开展了广播电视和网络视听区块链技术应用研究工作，编制了广播电视和网络视听区块链技术应用系列白皮书，从总体应用、内容审核、县级融媒体、内容版权管理、监测监管等方面指导相关领域区块链技术创新应用。

本白皮书为系列白皮书的内容审核篇，面向内容审核规范现状和技术现状，深入分析了媒体深度融合时期内容审核面临的痛点，结合区块链技术优势提出了基于区块链的内容审核方案和主要内容审核业务流程，并分析了该方案在媒体内容全流程管理方面的应用扩展。

本白皮书指导单位：国家广播电视总局科技司

本白皮书主要起草单位：国家广播电视总局广播电视科学研究院、深圳市迅雷网络技术有限公司、北京交通大学、北京长亭科技有限公司、北京新创智链科技有限公司。

本白皮书主要编写指导：孙苏川、关丽霞

本白皮书主要起草人：施玉海、张伟、牛妍华、陈哲、何晶、龚媛嘉、张骁、郜帅、张景驰、王桥、朱琳、李玉刚、孙海波、王晓光、徐杰、刘杰、叶凡、刘晶磊。

目 录

1. 概述

内容审核是各内容播出机构依据国家和行业相关法律法规、自律公约等对节目内容进行审核把关，通过识别节目包含的文字、图片、音频、视频中的涉黄、涉暴涉恐、敏感、虚假信息、违规广告等内容，来决定其是否适宜在相应平台上传播。依据“谁播出谁负责”的政策，传统的内容审核通常由内容播出机构主导，基于自有内容审核团队和审核平台开展内容审核，以保障自身平台内容的合规性。但是，随着媒体业务的持续繁荣发展，内容播出机构难以应对海量内容的审核需求，亟待建立高效的内容审核机制。

本白皮书面向符合“先审后播”规定的广播电视节目和网络视听节目[①]，以提升内容审核效率、改善播出机构审核压力为主要目标，并且引入区块链、人工智能等新兴技术，从技术角度解决内容审核当前所面临的困境，通过整合内容审核资源，为内容播出机构提供高效可靠的内容审核第三方服务[②]，大大降低内容播出机构的审核压力。同时，为监管机构提供必要的内容审核业务监管接口。本白皮书适用于符合“先审后播”规定的内容审核系统设计、审核流程制定和审核监管溯源等方面。

此外，内容审核区块链系统在设计、开发、建设、实施过程中，应同步考虑内容分级审核、主体责任落实、监管需求等问题。

第一，面向多样化的媒体内容，应根据内容价值、审核时限、播放范围等不同需求采取差异化分级处理机制。例如：对电视剧等高价值内容采用加密存储、安全下载等保护机制，播出机构具有是否采信第三方审核结果的权限，进一步保障内容审核质量。对海量的 UGC 短视频内容采用单次审核方式提升审核效率，同时避免加密等中间处理环节。

第二，压实内容审核区块链各参与方的主体责任。为保障内容审核区块链系统的良性运转，在建设的同期应明确内容提供机构、内容审核机构、内容播出机构及管理机构的主体责任，同时建议对所有信息进行自检存证。管理机构同步更新相关法规。

[①]本白皮书所述方案不适用于网络直播业务。

[②]内容审核由内容播出机构主导，根据内容审核业务量、内容审核时限等需求自行决定是否采用第三方审核方式。本白皮书所述方案属于第三方审核方式，基于区块链的信用机制为内容播出机构提供可信的第三方内容审核服务。内容播出机构的内部审核不在本白皮书讨论范围内。

第三，内容审核区块链为监管机构提供必要的、可扩展的内容审核业务监管接口。支持监管机构和平台运营方的主动监管和被动监管，提供内容审核交易记录的查询、管理和追溯能力，并定期提供审核业务报告，以保障内容审核区块链的良性运转。

2.内容审核现状及需求分析

内容审核是依据国家和行业相关法律法规、自律公约等对广播电视节目和网络视听节目进行审核把关，通过识别节目包含的文字、图片、音频、视频中的涉黄、涉暴涉恐、敏感、虚假信息、违规广告等内容，来决定其是否适宜在相应平台上传播。对于适合传播的内容进行确认，确保其能够顺畅传播，并得到推荐；对于不适宜向特定用户（如未成年人）传播的内容进行甄别和处理，防止其传播扩散；对于违反法律法规和有害的内容，则严禁播出并追究上传者和上传单位的责任，从而达到净化网络空间的目的。

2.1 内容审核发展历程

广播电视传统媒体发展初期，主要采用人工审核的方式进行内容审核，内容制作机构、电视台、有线网络公司等播出机构都有各自的内容审核团队，依据国家、行业的相关标准和规范，由专业人员对媒体内容进行监看审核。严格执行复核复审、重播重审、播前审核等制度，同时为了避免人工审核的漏审错审，采用多级审核人员严格把关，审核质量较高。

随着媒体业务的发展，人工审核方式呈现出效率较低的问题。受益于信息技术特别是人工智能技术的快速应用，机器审核技术逐渐成为内容审核的重要辅助手段，广泛应用于互联网内容服务平台。机器审核主要采用关键词过滤技术、图像识别、音频识别等智能审核技术。

（1）关键词过滤

关键词过滤是文本内容审核的有效手段，平台通过预先设置关键词或敏感词审核过滤文本内容，技术上容易实现，成本也比较低。例如：Facebook 在 2009 年创建了第一个内容审查的关键词，共包含有 15,000 个词。其他社交媒体也都设置有各自的“关键词”，并且在不断更新。

（2）智能审核

智能审核采用语音识别、图像识别、人工智能等技术，基于大数据的海量样本，从图片、文字、视频三方面把关，智能识别违规内容是媒体内容审核的重要手段。例如：2015 年 3 月，Twitter 推出了“Quality filter”功能，目的是删除包含威胁、攻击性或者虐待的语言、重复的内容，以及从可疑账号发送的信息。

2017 年，由 Facebook、YouTube、Twitter 和微软合作成立的反恐怖主义全球互联网论坛（GIFCT）采用人工智能识别和共享数据库技术有效地过滤，并且删除了大量的恐怖主义视频和图片。2017 年 3 月，Instagram 使用照片模糊的方法处理含有侵犯性内容的照片[①]。

当前，互联网视听平台逐渐发展成熟，UGC 内容呈现井喷式增长，媒体产业进入繁荣发展阶段。与此同时，恶意用户的反过滤手段也持续影响内容审核的质量。虽然机器审核技术可以在色情识别、暴恐审核、涉政敏感、违禁品检测等方面有不错的审核效果，但是在意识形态把关方面仍然存在缺陷，因此，人工审核仍是不可或缺的把关方式。2014 年，Facebook 和 YouTube 扩大内容审核团队以应对海量内容审核的需求；2020 年 4 月，Netflix 推出可供家长使用的内容过滤工具，可以按照标题删除电视或电影，将集中式的内容审核压力分散给用户。人民网计划在2021年将内容审核团队扩增到3000人，以应对庞大的内容审核压力。

内容审核需求的快速增长催生了一批第三方审核云机构，例如：网易云盾、七牛云、阿里云、百度云、华为云等都面向各类内容服务平台和用户，提供机器审核和人工审核相结合的内容审核服务。用户只需将被审核内容上传至云端，就可以等待审核结果。第三方审核云服务使得视听服务平台无需建立庞大的审核团队，也不需要购买和维护复杂的机器审核系统，可以较大程度降低人工成本和运维成本。但是由于内容安全责任的主体仍然是视听服务平台，第三方内容审核云缺乏必要的评价约束机制来保证审核质量。

2.2 内容审核规范现状

本白皮书面向广播电视节目和网络视听节目，由于这两类节目的播出机构不同、播出渠道不同、用户终端也不同，内容审核的标准和流程也存在一定程度的差异。广播电视节目的播出机构主要包括各级电视台、各级有线电视网络公司，网络视听节目的播出机构包括 IPTV 播控平台、互联网内容平台等。各类机构严格遵循“谁播出谁负责”的播出负责制，并制定相应的内容审核规范，确保安全播出。

[①] 基于区块链的内容信息审查研究[J]，海峡科技与产业，2019。

2.2.1 广播电视节目内容审核现状

广播电视节目的审核面向各类待分发或播出内容的采集、编辑、审核、编排和分发等全过程。所有引进的内容必须执行严格的审核程序方可进行分发或播出。内容审核对象包括自拍（制）、合作、购买、自购、收录、交换、受赠、征集、下载等形式的内容素材。内容审核包括两方面：一方面从播出角度审核内容本身的质量是否符合播出的技术标准。另一方面是从思想性、政治性和观赏性三方面把握平衡，考查审核内容是否符合社会舆论和道德规范等国家法律规范或行业标准。

内容审核根据业务发展需要采用灵活和安全的多级审核方式，在内容政审等关键环节执行复核复审、重播重审、播前审核等制度。可采取的审核形式包括：

（1）内容引进审核。内容在引进阶段的审核，称为引进审核。审核人员根据审核标准对新引进内容进行初步的审核。

（2）初审。内容第一次播出之前的初次审核，称为初审。审核人员依据审核标准对内容进行详细的审核。对审核规范中规定的应当删减、修改的内容进行剪辑。在保证内容安全、连贯的前提下，兼顾思想性和观赏性。

（3）复审。内容初审后的第二次审核，称为复审。逐一对照历史和新时期标准对内容进行审核。

（4）终审。对复审完成的内容进行抽样检查，称为终审。再次对照历史和新时期标准对内容进行审核。

（5）播出审核。内容播出前对其进行的审核，称为播出审核。根据数字电视前端系统和终端设备情况，对内容进行播出的适应性技术审核。

（6）重播重审。对已完成上述审核程序并已存储的内容，如需重播，播出前的审核称为重播重审。逐一对照历史和新时期标准，特别是结合当前新形势，对内容进行再次审核。

2.2.2 网络视听节目内容审核现状

根据中国互联网络信息中心（CNNIC）第 45 次《中国互联网络发展状况统计报告》的统计[①]，截至 2020 年 3 月，我国网络视频用户规模达 8.50 亿，较 2018 年底增长 1.26 亿，占网民整体的 94.1%；其中短视频用户规模为 7.73 亿，较 2018

[①]中国互联网络信息中心（CNNIC），第 45 次《中国互联网络发展状况统计报告》，2020 年 4 月。

年底增长 1.25 亿，占网民整体的 85.6%。网络视听用户规模快速增长，已经成为仅次于即时通信的第二大互联网应用类型。

中宣部副部长，国家广播电视总局党组书记、局长聂辰席多次强调，网络视听业务要努力做主流思想文化建设者、做优质精神食粮提供者、做新业态新服务开拓者、做清朗网络空间维护者。特别强调网上网下统一导向、统一标准、统一尺度，网络视听节目服务机构要牢记文化责任和社会担当，坚持把社会效益放在首位，切实落实主体责任，坚持依法依规运营，坚持先审后播，健全节目审核和播出流程，配强审核力量，把好导向关、内容关、质量关、传播关、人员关，加强对内容质量、功能格式、制作质量、版权关系等方面的监督审查，积极开展网络执法监管，探索建立版权工作长效机制，不断夯实行业健康、稳定、持续发展的基础。

为规范网络视听节目内容，广电总局近年来发布了一系列内容管理规定，包括《互联网等信息网络传播视听节目管理办法》（广电总局 39 号令）、《互联网视听节目服务管理规定》（广电总局 56 号令）、《关于进一步加强网络剧、微电影等网络视听节目管理的通知》、《专网及定向传播视听节目服务管理规定》（国家新闻出版广电总局令第 6 号）、《国家新闻出版广电总局办公厅关于进一步规范网络视听节目传播秩序的通知》（新广电办发〔2018〕21 号）等。中国网络视听节目协会也先后发布了《网络视听节目内容审核通则》、《网络短视频平台管理规范》、《网络短视频内容审核标准细则》等，进一步指导各网络视听节目机构开展网络视听节目内容审核工作，促进网络视听节目行业健康发展。

《网络视听节目内容审核通则》由中国网络视听节目服务协会于 2017 年 6 月 30 日发布，该通则将网络视听节目界定为：（一）网络剧、微电影、网络电影、影视类动画片、纪录片；（二）文艺、娱乐、科技、财经、体育、教育等专业类网络视听节目；（三）其他网络原创视听节目。要求互联网视听节目服务相关单位在网络视听节目内容审核方面，坚持先审后播和审核到位原则，要求审核员应完整审看包括片头片尾在内的全部内容，不得快进和遗漏。该通则还制定了详细的网络视听节目的内容导向、内容审核标准。

（1）专网及定向传播节目内容审核现状

为规范专网及定向传播视听节目服务秩序，促进行业健康有序发展，2016

年 4 月 25 日，国家新闻出版广电总局发布了《专网及定向传播视听节目服务管理规定》（国家新闻出版广电总局令第 6 号）。规范对象为：以电视机、各类手持电子设备等为接收终端，通过局域网络及利用互联网架设虚拟专网或者以互联网等信息网络为定向传输通道，向公众定向提供广播电视节目等视听节目服务活动，包括 IPTV、专网手机电视、互联网电视等形式。对内容提出不得违反法律、不得危害国家安全、不得破坏民族团结、不得危害社会公德等明确要求。同时规定，内容提供服务单位应当建立健全节目审查、安全播出等节目内容管理制度，配备专业节目审查人员。

以 IPTV 内容审核流程为例，内容审核人员包括编辑、负责人、主管及相关专审人员等，在节目整合、制作、审查等不同环节都要求注意内容的真实性、导向问题，重大、敏感和突发事件报道问题，版权问题等。IPTV 的内容审核遵从严格的三级审片制度，实行“先审后发”、“重播重审”、“再播再审”的审核要求。进行细致的责任分工，实行编辑自审、责编审发、主管审查制，使各把关环节切实负起责任，减少一般性差错，杜绝任何严重差错。

a.编辑自审是对内容进行基本校对，判断节目的内容版权、导向，避免侵权、失实和误导，杜绝错别字和内容缺失等。经频道负责人审阅认可后签发。若未经授权，编辑不得直接发布节目信息。

b.责编审发是对编辑制作的节目负有发布前的审查职责。编辑制作的节目内容必须经由频道负责人审阅认可后签发。编辑、策划制作的专题（栏目），也须由频道负责人审核发布。频道负责人根据对节目价值的判断，挑选重要节目信息呈现在频道（栏目）首页、推荐位等重要位置。

c.主管审查是对于在首页首屏呈现的重要节目信息进行严格审查把关。对把握不准的内容应及时上报、送审，经获准后方可发布。主管是团队节目信息传播安全的第一责任人，对团队成员采制、发布的节目内容，负有事先把关、事后监控的管理责任。

此外，还包括终端审核。从用户体验的角度出发，通过终端对上线节目进行终端审核，成立虚拟审核团队，对各业务地区的终端进行日日审核，包含直播、回看、点播及首页 EPG 审核。直播审核主要确认直播频道播放是否正常、直播页面的推荐位图片是否正常呈现、节目链接是否正确、编目信息文字及描述是否正

确、节目在点播中是否能正常呈现；回看审核是审核回看节目单是否完整，节目单与节目是否匹配，能否正常播放；点播审核是从点播架构、节目海报及编目信息是否正确、节目播放是否正常几个维度进行审核；首页 EPG 审核主要从 EPG 海报是否正确并符合当前报道氛围、海报链接内容是否正确、节目呈现及跑马灯是否正常等方面进行。

（2）网络短视频内容审核现状

随着短视频业务的飞速发展，海量的短视频内容也成为内容审核的重要工作。为提升短视频内容质量，遏制错误虚假有害内容传播蔓延，2019 年 1 月 9 日，中国网络视听节目服务协会发布了《网络短视频平台管理规范》（以下简称《规范》）和《网络短视频内容审核标准细则》（以下简称《细则》）。

《规范》规定了网络短视频平台实行节目内容先审后播的制度，平台上播出的所有短视频均应经内容审核后方可播出，审核对象涵盖节目的标题、简介、弹幕、评论等内容。同时，网络平台开展短视频服务，应当根据其业务规模，同步建立政治素质高、业务能力强的审核员队伍。审核员应当经过省级以上广电管理部门组织的培训，审核员数量与上传和播出的短视频条数应当相匹配。原则上，审核员人数应当在本平台每天新增播出短视频条数的千分之一以上。

《细则》制定了网络短视频内容审核具体内容，包括攻击我国政治制度、法律制度，分裂国家的内容，损害国家形象的内容，损害革命领袖、英雄烈士形象的内容，泄露国家秘密的内容，破坏社会稳定的内容，损害民族与地域团结的内容，违背国家宗教政策的内容，传播恐怖主义的内容，歪曲贬低民族优秀文化传统的内容等。

2.3 内容审核技术现状

目前，内容审核技术主要基于人工智能、大数据等机器审核技术，实现色情识别、暴恐审核、涉政敏感、违禁品检测等多个维度的审核，一定程度上降低人力资源成本消耗。

人工智能应用于内容审核的主要技术包括：语音识别、图像识别、图像分类。

（1）语音识别

语音识别的方法有三种：基于声道模型和语音知识的方法、模板匹配的方法以及利用人工神经网络的方法。其中，模板匹配的方法比较成熟，目前已达到实

用阶段。模板匹配方法包括四个步骤：特征提取、模板训练、模板分类和判决。隐马尔可夫法(HMM 方法)现已成为语音识别的主流技术，目前大多数大词汇量、连续语音的非特定人语音识别系统都是基于 HMM 模型的。

（2）图像识别

图像识别技术是立体视觉、运动分析、数据融合等实用技术的基础，在导航、地图与地形配准、自然资源分析、天气预报、环境监测、生理病变研究等许多领域具有重要的应用价值。通过图像识别可以快速准确地识别审核图片内容。

（3）图像分类

图像分类根据各自在图像信息中所反映的不同特征，把不同类别的目标区分开来的图像处理方法。它利用计算机对图像进行定量分析，把图像或图像中的每个像元或区域划归为若干个类别中的某一种，以代替人的视觉判读。本系统可以通过基于色彩特征和纹理的图像分类相结合技术，对需要审核的影像资源进行分类，快速分类图像，提高系统智能识别速度。

基于以上人工智能技术可以实现文字、图像、视频、音频的机器审核。

（1）文字审核一般基于敏感词过滤，预先设定一批关键词库并对词组进行排列组合，这批词库又会根据敏感性进行分类。系统会阻止用户发布敏感词汇，或将用户发出来的含有敏感词的内容直接删除。对于某些敏感性较低的词汇，发出来不会立即删除，需要经过审核人员过目进行二次审核。

（2）图片审核主要基于深度学习图像识别技术，通过针对目标特征专门训练的素材库和识别模型来甄别存在的违规图片。基于人工智能技术可以分辨色情裸露场景和宝宝裸露、正常健身、艺术品场景，同时能够识别视频中的色情语音或文字内容，降低媒体内容涉黄风险。

（3）视频审核结合 MD5 信息值作为视频的“数字指纹”，建立不合规文件的指纹数据库，通过指纹比对确定内容是否合法，则能避免不合规文件的重复分享。视频审核同样是基于深度学习图像识别云，将视频截图，由机器审核每一张截图的安全性。视频审核可以鉴别以下内容：结合暴恐敏感内容智能分析技术精准识别暴力、血腥场景、暴恐人物等违禁内容，对涉及暴恐的视频、语音均能准确识别，降低应用涉暴涉恐的风险；识别视频中的敏感人物和敏感事件，同时规避敏感问题；识别视频中的各类违禁品，如器官变卖、毒品等，避免内容涉及违

禁品的风险等。

但是，机器审核技术仍然存在识别精确度不足、意识形态把关难度大等问题，因此，人工审核仍然是必不可少的，搭建高效的人机协同审核系统可以降低人力消耗，提升审核质量，是目前认可的审核方式。目前，第三方审核机构成为内容审核领域的重要形式，通过改进的人工智能审核技术结合专业的审核人员队伍为各类内容提供商提供审核服务。

2.4 内容审核系统的部署需求分析

内容审核是保障内容播出安全和正确导向的重要环节，是掌握舆论宣传的源头。国家、行业、各执行机构制定了明确的内容审核规范和制度，人工智能、大数据等技术的应用对内容审核提供了较好的辅助。但是，内容审核仍然面临以下困境：

第一，网络视听节目服务单位审核压力大、人力成本高。现行的内容审核政策及规范要求网络视听节目服务单位建立完善的内容审核制度和审核意见留存制度，并配备相应数量的审核队伍和内容审核设备。这将对网络视听节目服务单位产成较大的工作压力和高昂的人力成本，对审核人员自身来说，也极容易造成心理疲倦和精神压力。

第二，网络视听内容的迅速膨胀对内容审核工作造成较大的压力。随着互联网技术的发展和大量用户的参与，网络视听节目尤其是短视频等业务面临海量内容且良莠不齐的问题，传统的内容审核方式效率低，无法应对网络时代视听节目内容产量高、传播快等特点。

第三，基于人工智能、大数据技术的机器审核还存在技术不够成熟，对意识形态和舆论导向把关能力欠缺等问题。当前对网络视听内容的监管和业务审核规则更新较快，适配难度比较大。机器审核需要根据变化的审核规则而频繁地更改代码、调整审核策略会使得整体操作成本变高，与此同时，针对不同的规则来制定模型也需要大量的训练数据，从现实角度来看存在成本高和难度大等问题。再者，由于各内容审核机构各自为战，虽然各有所长，但缺乏交流共享，审核经验无法得到有效的相互借鉴及经验提炼，导致人工智能训练素材标准性及完备性都还有较大的提升空间。

第四，审核人员素质参差不齐，不同地区、不同平台的审核标准不完全一致。

《国家广播电视总局关于进一步加强广播电视和网络视听文艺节目管理的通知》（广电发〔2018〕60 号）要求坚持同一标准、同一尺度，维护广播电视与网络视听节目的健康有序发展。网上与网下要坚持统筹管理、统一标准。由于审核人员培训难度大，掌握全面审核知识的审核员数量少，非主流、亚文化甚至扭曲价值观的内容元素夹杂于节目中传播，增加审核难度，因此，优秀的审核人员仍然是内容审核的重要因素。

第五，各审核机构、播出机构难以建立互信、公平、共享的协同机制，审核效率较低与内容繁荣发展的矛盾已然非常突出。目前，各类播出机构为确保安全播出，严格遵循“谁播出谁负责”的播出负责制，按要求建立 50～5000 人的内容审核团队，实现全部内容的全量审核，审核工作量巨大，同时针对其他来源的内容，不论是否已通过审核都需要进行重复审核，存在大量的重复审核工作，对人力物力形成巨大消耗。

综上所述，各内容播出机构的内容审核工作量与日俱增，而基于大数据及人工智能等先进技术的机器审核方式仍存在难以规避的短板，人工审核仍然是不可或缺的内容把关手段。为了突破困境，助力媒体融合良性发展，内容审核系统应在系统设计、部署及运营发展过程中具备以下能力：

第一，内容审核系统需具备整合现有内容审核资源的能力。内容审核资源包括第三方审核机构、现有播出机构的内容审核部门等，涉及专业的内容审核人员和机器审核系统。有效的资源整合可以提升机器审核的质量，降低播出机构的审核压力。本白皮书从行业引领的高度建立内容审核区块链，可以吸收专业的审核资源，培养高质量审核团队，为媒体内容良性发展提供保障。

第二，内容审核系统需具备高可靠的内容传输机制和可信的信用机制。相同内容的重复审核是当前内容审核的最大痛点，理论上，经过高信用等级团队审核的内容在一段有效时间内对播出机构是可信的。也就是说，基于安全传输机制保障内容的完整性，基于信用机制保障内容审核的质量，可以有效减少重复内容审核工作。本白皮书正是借助于区块链技术提供内容完整性验证和机构的信用机制。

第三，内容审核系统需具备智能审核系统的持续优化能力。面向人工智能技术日新月异的发展，以及监管部门政策和业务审核规范的频繁更新，内容审核系统应依托于稳定的研发团队和运维团队，持续跟踪技术发展，不断完善系统功能，

及时更新审核规则，以避免审核缺漏，并有效提升机器审核质量。

第四，内容审核系统需具备统一导向、统一尺度的审核标准制定能力。针对现有审核人员素质参差不齐的问题，内容审核系统在部署前和运营期间需要定期组织开展内容审核人员培训，提升人员素质，培养高质量的专业审核团队。针对审核标准不一致的问题，通过在内容审核系统中写入统一的审核规则，以及在规则更新时及时或同步在系统中进行变更，从人员和技术两方面保证提供统一高质量的审核服务。

综上所述，本白皮书提出了基于区块链的内容审核解决方案，充分运用大数据、人工智能等先进技术，结合区块链技术，整合具有审批资质机构的专业内容审核资源，基于统一导向、统一尺度的内容审核标准规范，为广播电视机构和各类网络视听服务单位及其他内容审核需求方提供高效可靠的内容审核方案，有助于加强对内容传播的安全管控，进一步引导和规范网络内容传播秩序，为打造良好的视听媒体内容生态环境提供基础保障。

2.5 区块链应用于内容审核的优势

区块链作为分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式，在内容传播领域具有广泛的应用前景。依托节点之间达成的分布式共识和可追溯、不可篡改的数据存储，可以将内容传播链条上的内容发布、内容加工、内容传播、内容审核等所有行为和数据进行完整记录，实现内容资源全程管理；同时结合 AI、大数据等技术的针对性应用，实现精准内容传播，保障有据可查、有迹可循。

区块链在内容审核的价值体现在自动流转、审核溯源、责任落实等方面。通过区块链技术生成内容标识的唯一性和内容审核的共识性，能够解决内容及其审核历史流转溯源可信的问题，通过区块链网络中多参与节点共同对内容及审核进行记录，可互相验证内容审核有效性并达成一致。借助区块链技术建设内容审核系统具有以下优势：

第一，有助于打通内容审核孤岛，构建各媒体平台间的互信机制。

各播出机构各自对内容进行审核，同一内容在不同播出机构需要重复审核，从而造成大量的资源消耗。基于区块链技术的内容防篡改，可以结合时间戳，保障内容的审核互信，促进内容审核效率的提升，减少违规内容重复审核工作量，

推动网络视听良性发展。

第二，有助于提升媒体内容传播的可信度。

通过基于区块链的内容审核平台，内容发布机构经由区块链平台进行验证，只有足够数量的节点统一才允许发布。区块链技术具有分布式账本、时间戳、篡改难度大、防欺诈等特性，可追溯到内容发布者，确保媒体发布内容的可信度。

第三，有助于构建内容审核机构的信誉评估及约束机制。

基于审核结果信息上链后可在联盟内公开、可信度高及可追溯等特性，内容审核机构的审核失误率和信誉度将公开透明且可追溯，结合基于审核贡献值的正向激励，可以有效推动各平台间审核人员进行审核经验及审核素材的共享，进而逐步建立对内容审核机构的评价机制，打造高质量的内容审核服务链。

第四，支持内容审核的安全监管溯源。

基于区块链数据的不可篡改性，监管机构可以定期对历史记录进行分析，并且根据监管结果形成审查平台安全态势分析，构建内容审核链安全监测平台。一方面，可以对恶性审核行为和安全风险进行监测预警；另一方面，基于链上的信誉记录机制，促进内容提供商持续制作和提供健康优质的内容。

综上所述，将区块链技术应用于内容审核，可进一步构建可信度高的媒体内容共享交换机制，提升内容审核效率，减少违规内容重复审核工作，促进内容制作生态的健康发展，提供内容审核、认证确权、隐私保护、可信交换和共享等应用服务，从而推进整个媒体内容的治理升级。

3. 基于区块链的内容审核方案

基于区块链的内容审核方案是依托区块链技术支撑的全流程防篡改、可追溯数据链条，结合人工智能、大数据等智能审核能力，打造以区块链为基础的分布式内容审核体系，为内容服务商、审核机构、播出机构搭建一个效率高、可信赖的集智能审核、追溯、认证、评分于一体的内容审核工作平台，提升审核效率、降低审核成本。

3.1 方案概述

基于区块链的内容审核方案是以目前已有的内容审核平台为基础，建设内容审核区块链网络，通过联盟链的方式将内容服务商、审核机构、电视台/播出机构和国家级审核监管部门联合起来。通过区块链特有的共识机制，在机构间实时同步内容审核结果和审核标准，实现审核链条在联盟体系内的公开透明，将不同机构间的信息流通阻碍降至最低，从而实现以最终审核效果为导向的可追溯、可定位的高效透明审核机制。基于区块链的内容审核方案架构如图 3-1 所示。

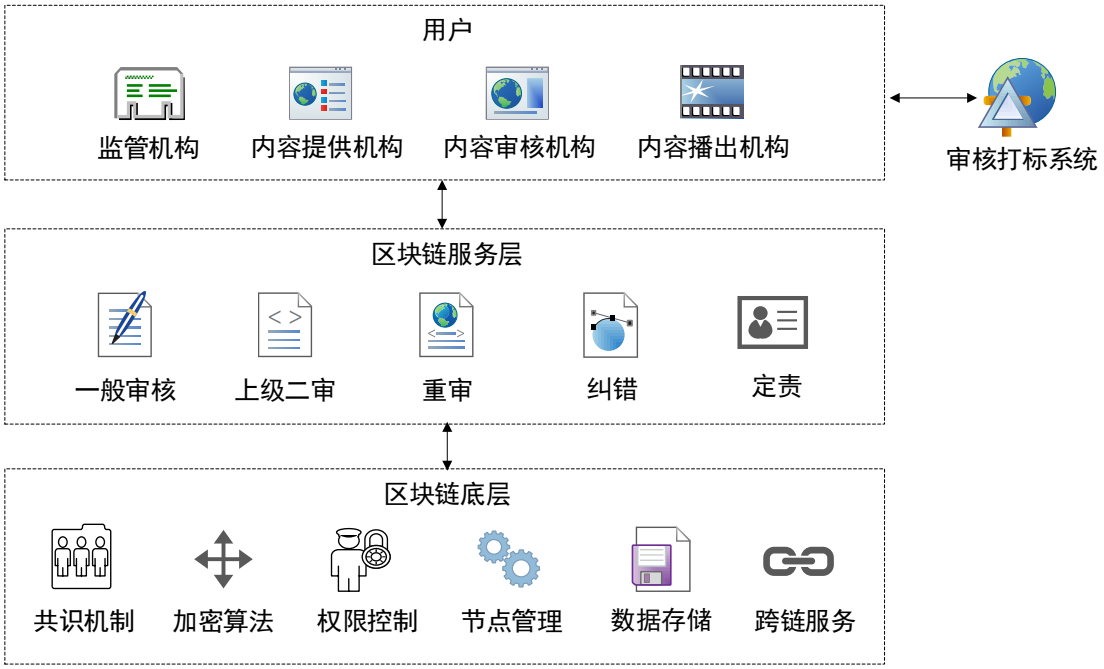

图 3-1 基于区块链的内容审核方案架构

内容审核区块链由区块链底层、区块链服务层和用户构成。用户包含监管机构、内容提供机构（或内容服务商）、内容审核机构、内容播出机构（或内容发布机构）四类，用户可根据业务需求调用区块链服务层的一般审核、短视频审核、

上级二审、重审、纠错、定责等功能组件。区块链底层包含共识机制、加密算法、权限控制、节点管理、数据存储和跨链服务等功能。底层节点逻辑与上层用户逻辑分离，区块链底层节点的所有者可以是用户角色中的任意一种，具体选择规则由相关用户能力或政策规则需求决定。

审核打标系统是基于现有审核机构进行功能扩展的系统。内容审核机构将内容审核结果信息发送给内容审核区块链，内容审核区块链再对信息进行处理。基于哈希赋予的文件唯一性标识和可追溯的公开审核记录，能够减少不必要的重复审核，并定位审核问题的环节，实现不可篡改、可追溯、高效率的内容审核服务。

3.2 内容审核区块链

内容审核区块链通过存证接口将内容审核记录上链，并通过底层节点的共识将包含该记录的交易打包进区块，通过区块链的 P2P 网络、密码学算法、时间戳等技术，保证了内容及审核结果的公开、透明和不可篡改。

内容审核区块链以适当的手段激励各个节点参与区块链共识。业务发生时数据多方同时确认并提供不可篡改记录。内容审核区块链可以极大地提升内容的审核效率和可信度，保障内容流转及内容审核信息历史追溯。可以克服传统模式存在的缺陷，任何节点都可以记录审核信任授权信息，且不可更改。

当需要提取内容流转及审查记录时，可以依据内容标识提取不同时间段的内容流转审核记录历史信息，这些信息都是基于区块链技术的共识以及验证存储和记录，在可信度和可靠性方面得到了区块链各参与方的共同验证，可以将可信的内容审核历史记录发布到应用平台，支撑安全监管。

3.2.1 区块链类型

区块链按照部署形式分为公有链、联盟链和私有链。其中，公有链是一种完全去中心化的区块链，所有节点都可匿名随时进出区块链网络，这对于内容审核业务来说是不利于监管和追责的。私有链适合机构或组织内部使用，适用于内部数据的访问和权限管理。联盟链既具有公有链“去中心化”的特点，又保留了私有链的“隐私性”，是一种部分“去中心化”的区块链。联盟链在组织机构间数据交互方面具有很大优势，相比公有链，联盟链节点数量较少，使系统运行效率提高，并且只有被授权的节点才能加入联盟链网络，有助于监管和追责；相比私

有链，联盟链由各节点共同维护，可信度高，同时具备较高的扩展性。因此，内容审核区块链以联盟链作为基础架构。

基于联盟链的节点准入机制，联盟链的所有节点身份同各参与方机构相互对应，保障节点行为所属机构可追溯。同时，基于数字身份和权限控制机制，写入区块链的数据都包含各参与方的数字签名，每一条审核信息任何人无法篡改且随时可追溯。因此，更适合于内容审核领域，一旦发现数据真实性问题，相关参与方都无法抵赖，造假操作会对其诚信造成恶劣影响，甚至要负法律责任。

3.2.2 区块链参与方

内容审核区块链参与方包括内容提供机构、内容审核机构、内容播出机构（或内容发布机构）和监管机构。

（1）内容提供机构

是内容发行机构或内容制作机构，为内容播出机构提供内容源。

（2）内容审核机构

由政府部门组织确定并给予权威审查授权的审核节点，传统内容播出机构的内容审核部门通过审查后，也可授权作为内容审核机构节点加入内容审核区块链。

（3）内容播出机构

包括传统的广播电视播出机构（例如：各级电视台、各级有线电视网络公司等）、新媒体平台（例如：各级 IPTV 播控平台、OTT 平台等）以及其他网络视听节目服务单位（例如：爱奇艺、优酷、腾讯等互联网视频服务平台）。

（4）监管机构

是具有公信力的监管部门，具有审核联盟成员的权利，监管部门保存全量数据，并拥有查看每一笔交易以及参与者管理的权限。

在多方参与的联盟链中，各参与方在分工协作的同时要职责分明、各司其职。各参与各方根据用户角色具有不同的权限，包括各类业务数据的写入、读取、业务权限控制，由区块链数字身份系统实现，底层链节点负责对这些权限进行验证、响应和处理。

3.2.3 区块链节点

内容审核区块链节点分为记账节点、普通节点和轻节点三种类型。

记账节点负责对各用户的交易信息，如申请、审核、智能合约调用等进行排

序、打包和共识。普通节点负责同步并维护完整的账本数据。轻节点负责对与其相关的区块及交易信息进行验证，并且通过建立连接的记账节点或普通节点进行交易信息查询。

在内容审核区块链中，建议监管机构与内容审核机构至少建设一个记账节点和一个普通节点；内容播出机构可以建设普通节点，也可以根据需求建设轻节点。

3.3 审核打标系统

根据内容审核现状调研，现有的内容审核机构通过基于人工智能的机器审核方式，基本可以实现对色情、暴恐、涉政敏感、违禁品检测等内容进行智能识别。本白皮书基于现有内容审核平台进行建设，不对内容审核平台的技术细节进行描述。

审核打标系统可基于现有内容审核机构的审核平台进行必要的功能升级或模块开发，使其达到以下功能或管理要求：

（1）采用对现有平台进行升级或新建相关模块的方式，实现人工智能和大数据能力支持，同时应进行持续优化，以保障最大程度上提升机器审核效率、降低人力成本。

（2）增加内容标签打标功能模块，实现对审核对象标注必要的内容标签，可以提升后续重审的效率。

（3）明确人工审核的不可或缺性，注重人工审核协同操作模块的功能开发和维护。

审核打标系统具备文本审核、图像审核、视频审核和内容标签打标等功能，采用机器审核和人工审核相协同的方式，其功能模块如图 3-2 所示。

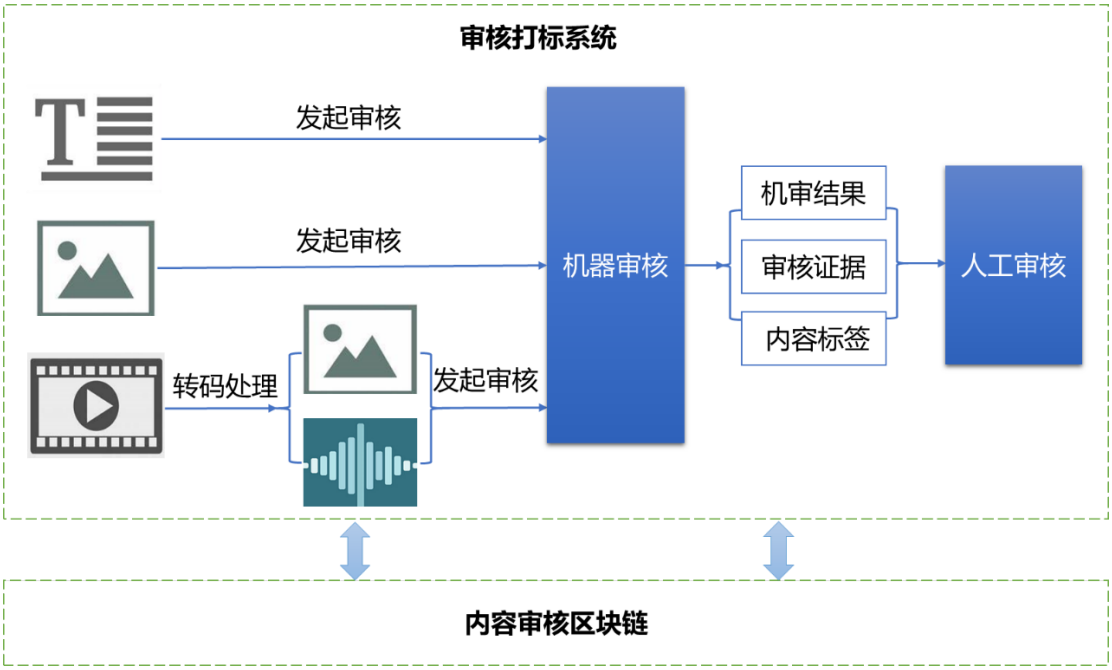

图 3-2 审核打标系统

对于视频内容，先进行抽帧转码处理，然后提取重要帧的图片和音频文件分别进行审核。机器审核完成后，将机审结果、审核证据和内容标签等转发给人工审核功能模块，完成复审工作。整个审核流程完成后，最终的审核结果、审核证据和内容标签将发送给内容审核区块链。

审核打标系统包括机器审核、人工审核和接口模块三部分。

（1）机器审核

a. 对审核对象中的视频按设定的频率进行截帧（如可以自行设定每秒截取 1-24 帧画面），对截取画面（含画面上的人物、特定对象、字幕、广告等）进行特征识别、分析、比对，自动根据敏感库、特征库及自我学习库中的数据判断是否包含可疑内容，并标注可疑原因、类型（涉黄、政、暴、恐、军、外等）和等级（如初步可疑、重点可疑、禁止播出）。

b. 对审核对象中的音频部分进行自动获取和识别，并进行对应的分析、比对和判断。

c. 在对审核对象进行特征识别的同时，采集审核对象的关键信息，并标注相应的内容标签，包括关键人物、时间、主要事件、关联人物、关联事件等，如表 3-1 所示。内容标签将被审核对象进行标签化，当需要重播重审、审核规则或策略发生变更时，内容标签可以迅速定位被审核对象，大大提升审核效率。

表 3-1 内容标签字段

字段	说明
person	关键人物
time	时间
event	事件描述
relperson	关联人物
relevent	关联事件

d.对审核对象进行分类和分级，对应审核人员的分类和分级，将不同类型和等级的内容匹配派单给相应类型和等级的审核人员进行复核。

（2）人工审核

a.基于智能审核的结果进行相应的审核操作。对于文字类的待复审内容，审核人员根据 AI 标记进行复核确认；对视频类的待复审内容，审核人员进行视频播放审核，重点审核关键帧（含标记图片），并进行必要的编辑和点评。

b.审核人员对审核对象的标签进行复核，所有完成审核的内容都建议添加内容标签。

c.审核人员对复核结果进行回填，进一步提升同类可疑内容提取和横向对比分析等能力。

（3）接口模块

内容审核工作完成后，审核打标系统将审核结果、审核证据和内容标签通过接口模块打包发送给内容审核区块链，链下审核工作完成。

4. 内容审核区块链架构

内容审核区块链架构由应用层、服务层、核心层、基础层和安全机制组成，如图 4-1 所示。

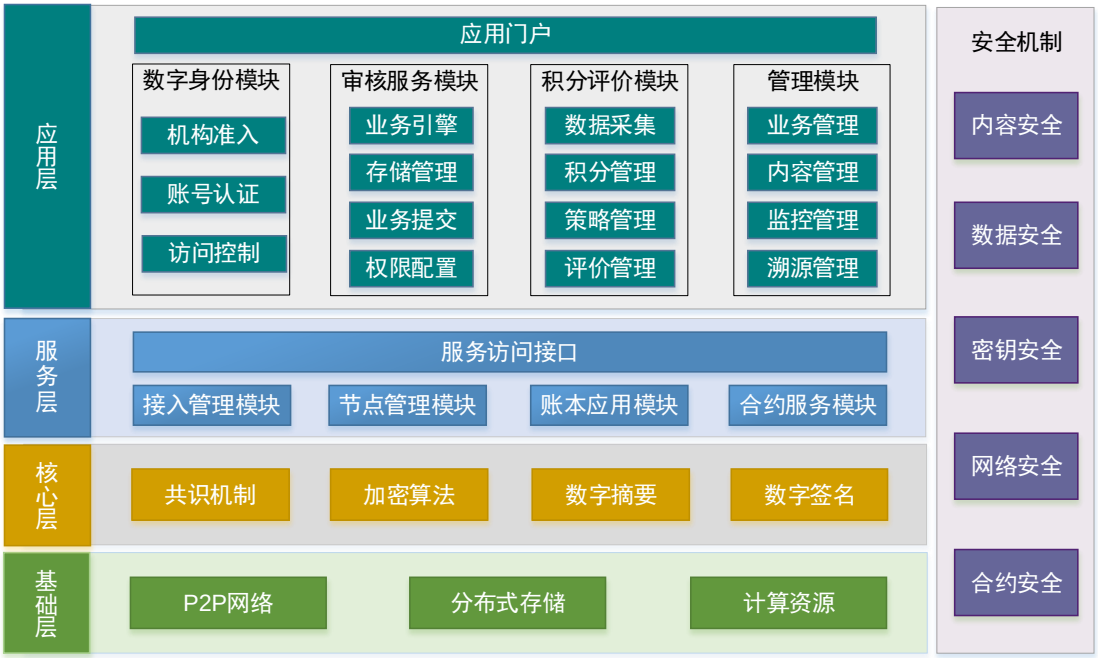

图 4-1 内容审核区块链架构

4.1 应用层

应用层是内容审核区块链系统面向用户的入口，为参与方访问区块链服务提供交互界面，并负责用户相关的管理服务。应用层由数字身份模块、审核服务模块、积分评价模块、管理模块和应用门户模块组成。

4.1.1 数字身份模块

区块链技术通过密码学及共识算法解决了链上数据的信任问题，其中交易真实性、数据隐私性及身份权限等问题可通过数字身份解决。数字身份模块负责内容审核区块链各参与方的身份管理和身份认证，确认用户对资源的访问权限，保障区块链访问控制策略的可靠运行。数字身份模块包括机构准入、账号认证和权限访问功能组件。

（1）机构准入

为保证内容审核区块链中的所有节点可追溯、节点行为所属机构可追溯，节点身份与实体机构一一对应，做到节点行为所属机构可追溯[①]。本方案为所有节点颁发 CA 证书，为各机构备案公钥证书到区块链，节点连接时通过验证颁发的 CA 机构证书有效性来判断节点是否准入，对已连接的节点行为进行追溯。联盟链管理员生成 CA 根证书并把 CA 根证书公钥提供给所有节点使用。CA 根证书为每个节点颁发 CA 用户证书或同一机构下所有节点使用同一 CA 机构证书。节点的证书都应该写入系统合约中，若不写入，相应的节点将不允许通信。证书准入状态采用表 4-1 字段进行说明。

表 4-1 证书准入状态字段说明

字段	说明
hash	公钥证书哈希
pubkey	公钥证书
orgname	机构名称
notbefore	证书生效时间
notafter	证书过期时间
status	证书状态
whitelist	IP 白名单
blacklist	IP 黑名单

（2）账号认证

账号认证功能组件通过数字签名实现对请求发起各类交易、存证以及他证等操作的授权。在内容审核区块链中，每个账户拥有一对公私钥对，其中公钥作为公开信息，生成的地址作为账号 ID 使用。私钥作为不公开信息，结合密码学及公开的公钥，用以判断是否真正持有该账号的管理权限。例如，当内容审核机构完成对某一内容的审核后，先用私钥对内容进行签名，然后上链。其他节点收到消息后，通过对数字签名进行验证完成对该内容审核机构的验证，以判断是否审核或是否播出。

（3）访问控制

[①]对于内容播出机构的内部审核部门，如果有提供第三方内容审核服务的意愿，在经过监管部门资质审核后也可作为独立的内容审核机构加入联盟链。

权限控制功能组件是区块链安全的重要部分，功能包括联盟用户身份管理、用户角色管理和权限配置。按访问或读取资源的权限不同定义多个角色，按角色授权不同用户访问或使用资源的权限及安全规则，使用户只能在有限范围内进行写入或读取数据资源，而在其读取范围外的数据则会进行加密保护。

内容审核区块链各参与方根据角色的不同，拥有的权限也不同，在有限的范围内写入和读取数据，而在其读取范围外的数据则会加密保护。联盟节点成员之间权限、职责、资源、合作形式都形成共识，在彼此信任、权限平等的平台上，实现内容资源共享、内容交易、资源置换、内容授权等业务。保证内容资源流转、收益分配流转等链条清晰明确。

a.权限控制模型

内容审核区块链建议使用 ARPI（Account-Role-Permission-Interface）权限控制模型[①]，设置账号、角色、权限、接口四类对象。

账号和角色的对应关系是 N:1，即一个账号只能对应一个角色，但一个角色可以包含零到多个账号。在实际操作中，同一角色可能有多个账号，但每个账号使用独立且拥有唯一的公私钥对，发起交易时使用私钥进行签名，接收方通过公钥进行验签，以确认交易由哪个账号发出，实现交易的可控及后续监管的追溯。

角色与权限的对应关系为 N:M，即同一角色拥有多个权限的集合，即同一权限也能分属于多个角色。权限粒度细化到合约的接口级别，即某一角色下的账号如拥有某个权限，则该账号能调用该权限下合约的一个或多个接口。

需要注意的是，如果某个机构同时兼顾多个角色的职责，这个机构的人员需要进行交易，又可以作为运维人员参与系统维护，那么应给该机构分配多个账号，每个账号的私钥掌握在不同的人员手里，每个人员只负责一种角色的工作，以规范操作流程。

b.角色和权限设计

内容审核区块链上的各参与方各司其职、分工合作，以避免越权操作带来的风险。面向角色的权限控制和合理的账号分配管理，使对应到不同角色的账号职责清晰，活动行为可控，容易追溯。角色设计如下：

①链管理员

[①]ARPI 是 FISCO BCOS 平台基于系统级权限控制和接口级权限控制的思想提出的权限控制模型。

由内容审核联盟链的各参与方共同选出一个委员会，一个或多个机构可获得链管理员权限，负责组织各机构加入联盟链，为各机构分配对应的机构管理员和交易员权限，以及管理链上应用的生命周期等。链管理员根据管理策略，依次给其他系统管理员账户分配节点部署权限、配置策略修改权限等。链管理员对联盟链的日常运营负责，一般不直接参与链上业务交易。

②系统管理员

实施联盟链部署和运维活动的人员，通常由各参与方机构的运维团队承担，或者由平台运营管理者统一组织运维管理。运维人员负责发布和管理应用，管理区块链的节点物理资源，启停节点服务，修改本地节点的系统配置参数，一般不会参与业务交易。链管理员可以根据约定的治理规则来分配系统管理员权限，例如：只允许指定的账户部署合约，设定合约部署权限，其他账户则不能随意部署合约。

③内容审核员

实施内容审核职能的人员或机构，可以由具有资质的第三方内容审核机构承担，或者由其他机构的内容审核部门或团队承担。内容审核员负责处理全链各类内容的审核服务需求，审核接入记录、审核执行记录、审核结果信息均作为区块信息写入区块链，实现审核基础信息在全链公开透明。

④交易用户

交易用户发起业务交易，用户向区块链发送业务交易请求，业务交易主要是调用合约和读写用户表，可以根据业务逻辑，结合用户表权限和合约接口权限来灵活控制。

⑤监管机构

监管机构负责制定业务规范、审查业务数据、监督管理联盟链的运行状态，维护服务的合法、稳健运行。监管机构保存全量数据，并拥有查看每一笔交易以及参与者管理的权限，一般不参与联盟链的日常运作管理。

⑥应用开发者

应用开发者负责将可发行的软件提交到平台的应用仓库，包括智能合约、APP等。在 DO 分离的管理模式中，开发者不直接对联盟链进行操作，由运维管理者进行软件发布、参数配置等操作。为了跟踪应用的使用情况，开发者有查看相关

应用统计数据的权限。

4.1.2 审核服务模块

（1）审核服务模块的主要功能组件

审核服务模块是内容审核服务的交互中介，一方面接入链下审核打标系统的审核结果信息，另一方面为内容审核各参与方提供业务的接入和提交，以及数据交换服务等功能。审核服务模块主要包括业务引擎、存储管理、业务提交和权限配置功能组件。

a.业务引擎

业务引擎功能组件负责向各参与机构提供内容审核业务接入服务，包括业务查询、业务接入、文件唯一 ID 分配、存储空间分配等功能。业务查询为各参与方提供业务进度查询、业务记录查询、内容文件查询等服务，可按照交易编号或区块高度进行查询。业务接入为各参与方提供内容审核服务请求接入，也为审核机构提供审核结果反馈接入。文件唯一 ID 分配负责为各参与方提供的文件分配全网唯一 ID。存储空间分配根据存储管理模块提供的信息为各参与方提供文件存储信息。

b.存储管理

存储管理功能组件用于内容文件库存储空间管理，为被审核内容文件、审核过程文件、不同审核版本文件等提供存储管理。

c.业务提交

业务提交功能组件负责将用户请求、用户结果等信息进行结构化处理，生成区块哈希值，使之可以与联盟链网络进行交互。具体来说，把数据打包为交易，通过 P2P 网络进行广播，最终被超级节点接受处理，完成“上链”操作。

d.权限配置

权限配置功能组件为相关内容文件配置访问权限，包括访问时间和访问方式等，只有授权用户才能够访问内容文件。

（2）审核服务模块的工作思路

a.内容审核服务请求方（如内容制作机构、内容播出机构等）发起内容审核请求，内容审核请求信息包括文件 ID、文件哈希、文件位置信息、请求业务类型、时间要求和审核请求机构签名。内容审核请求消息字段说明如表 4-2 所示。

表 4-2 内容审核请求信息

字段	说明
fileID	被审核文件唯一 ID
filehash	被审核文件哈希，用于文件完整性保护
fileloc	被审核文件存储位置索引
servtype	用于标识请求业务类型，例如：一般审核、短视频审核、上级二审、定责、纠错、重审等
timeline	完成时间要求
sig	内容审核请求机构签名

b. 内容审核服务提供方接收到请求后，到指定位置获取原始文件，并进行哈希运算以确定文件的完整性。

c. 内容审核服务提供方进行内容审核，并通过审核达标系统生成内容标签。

d. 内容审核服务方将审核结果信息执行上链操作。审核结果信息包括原始内容 ID、完成审核后的内容 ID、内容标签、当前文件哈希、当前文件位置信息、审核问题文件 ID、审核问题文件哈希、问题文件位置信息、审核员 ID、审核日期、审核机构签名等。内容审核结果信息字段说明如表 4-3 所示。

表 4-3 内容审核结果信息

字段	说明
fileID	被审核文件唯一 ID
nfileID	完成审核的文件 ID
conttag	内容标签，包含人物、事件、时间、关联人物、关联事件等信息
nfilehash	完成审核的文件哈希
nfileloc	完成审核的文件存储位置索引
errID	审核问题快照文件
errloc	审核问题快照文件存储位置索引
auditorID	审核人员 ID
time	审核完成时间

字段	说明
sig	审核机构签名

内容审核结果信息是内容审核区块链的核心存证信息，一旦完成上链，内容审核结果信息即作为内容审核证书永久保存，每个内容文件每经过一次内容审核，无论审核结果是否发生变化，都会生成一个内容审核证书，为后续审核、交易、溯源、评价等提供基础依据。

（3）审核服务模块的用途

鉴于以上机制，审核服务模块可以实现信息验证、信息溯源、重播重审和事故定责等用途。

a.信息验证

审核机构发布的每一条审核信息都包含由审核人账号的私钥结合相关信息（包括审核日期、内容 ID、审核形式等）生成的数字签名。应用层模块在处理信息时，通过审核人已公开的公钥信息对其合法性进行验证。

b.信息溯源

由于每一条审核信息都包含了账号及其数字签名信息，因此可以通过两者确定其相应的审核人信息或审核申请的交易用户信息。

c.重播重审

根据需要引用已完成审核程序的内容是包含在新审核信息中，审核人对内容进行重播重审，其中将包含审核人私钥结合相关信息生成的数字签名。

d.事故定责

对于播出事故中涉及的相关内容，通过其内容 ID 等标识，收集相关联的审核信息，并通过信息溯源模块追责相关审核人，并把相关信息打包于交易中，进行全网广播，用以后续积分评价模块的运行。

4.1.3 积分评价模块

内容审核区块链应建立完善的积分评价体系，不但能促进内容审核机构提升审核质量，还可以依据审核评分的结果形成专业化分工。随着积分评价体系的逐步完善，将推进内容审核领域建立一批高质量的审核机构，提供优质专业的审核服务。

（1）积分评价模块三个要点

a. 原始分

依据不同内容审核机构的资质、实力、背景和基本信息，每个内容审核机构都有一个原始分，作为评价其审核能力的初始参考。

b. 分数加减

根据内容审核机构在实际审核中的表现，对其审核能力进行加分或扣分操作。例如：自身出现一次审核失误按既定积分策略扣分，发现一次其他机构审核失误按既定积分策略加分。

c. 评价策略

对于内容审核结果是否存在失误、失误的严重程度如何，需要由权威机构制定公平合理的评价策略条目保障内容审核区块链的良好运行。

（2）积分评价模块功能组件

积分评价模块包括数据采集、积分管理、策略管理、评价策略库四个功能组件，其功能架构如图 4-2 所示。

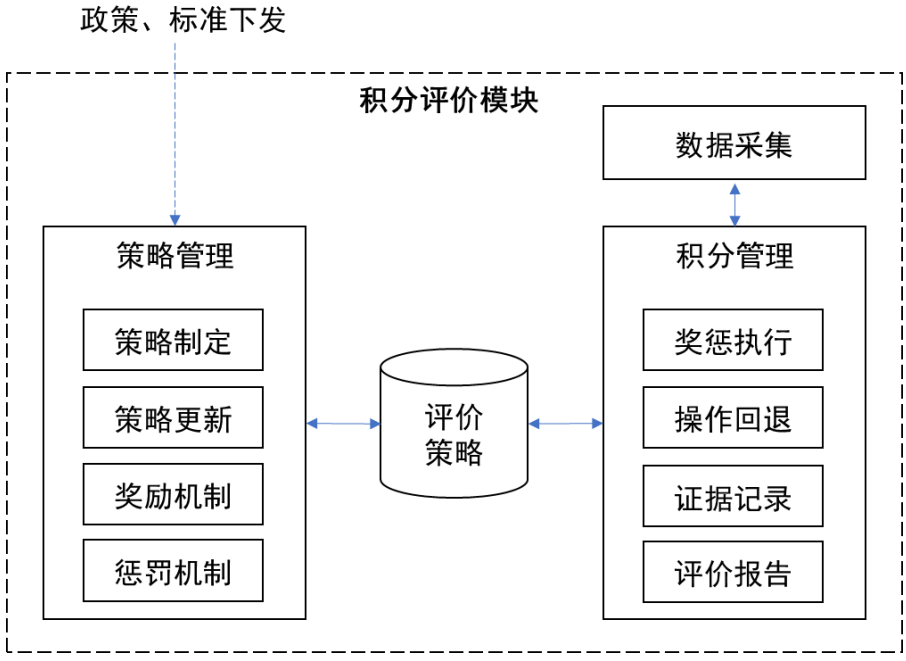

图 4-2 积分评价模块

a. 数据采集

数据采集功能组件负责统计内容审核机构的审核失误、审核校正、事故发生率等数据。

b. 积分管理

积分管理功能组件包括奖惩执行、操作回退、证据记录、评价报告等功能。

奖惩执行负责管理各内容审核机构的积分，依据评价策略进行加分、扣分、取消等操作。操作回退提供积分操作的回退功能。证据记录负责将证据信息打包于交易中进行全网广播，作为后续惩罚的凭证。评价报告提供按时段的积分评价报告。

c.策略管理

策略管理功能组件包括策略制定、策略更新、奖励机制、惩罚机制等功能。策略制定根据现有内容审核评价策略进行制定，包括积分增加、积分扣除等策略。策略更新根据新的政策、标准、规范、制度等进行策略修订。奖励机制用于鼓励内容审核区块链参与方积极参与共识，也鼓励内容审核机构提升审核质量，对一定时间内严格遵照审核标准，并且内容审核事故发生率在某一阈值下的审核人及其节点进行奖励。惩罚机制用于对交易行为异常、审核失误率超过阈值的机构实施进行惩罚。

d.评价策略

评价策略功能组件是评价策略条目的存储数据库。

积分评价体系需要随着内容审核区块链的运行逐步完善，经过稳定的运行，积分可以成为代表审核机构在内容审核领域专业度的指标，可以成为其他播出机构判定审核后是否需再审、是否可上线的重要依据，可以此反向督促内容审核机构提升审核质量，进一步降低重复审核的工作量，打造良性运转的内容审核环境。

4.1.4 管理模块

管理模块包括业务管理、内容管理、监控管理、溯源管理等功能组件。

（1）业务管理

业务管理功能组件负责对内容审核区块链相关的所有业务记录进行管理，包括业务请求、内容审核、重审、纠错等业务记录、交易记录、审核结果记录、文件播出（发布）记录等进行管理，并提供相应的业务查询接口。

（2）内容管理

内容管理功能组件提供所有文件内容的管理，包括原始文件、被审核过程文件、恶意内容文件、问题快照文件等管理，包括文件 ID、文件哈希、文件存储位置索引等信息，可通过区块头或内容 ID 进行检索。

（3）监控管理

监控管理功能组件负责区块链运行状态监控和故障监测，并提供事故和问题

报告的收集。

（4）溯源管理

溯源管理功能组件负责向监管机构提供内容审核记录查询溯源，以及问题的跟踪和报告服务。

4.1.5 应用门户模块

应用门户模块提供内容审核区块链业务访问和业务查询功能，由用户界面和事务提交两个模块组成。用户界面模块为内容审核区块链中的用户与区块链服务提供交互功能。事务提交模块可以实现将用户的事务请求（例如：内容审核交易申请、内容审核记录查询等）提交到区块链网络。

4.2 服务层

服务层为应用层提供可靠高效的区块链访问和监控功能。服务层通过调用核心层功能组件，提供统一的接入和节点管理服务。同时，还通过高效缓存、可靠存储、负载均衡等技术，提供可靠的区块链服务能力。服务层包括接入管理模块、节点管理模块、账本应用模块、合约服务模块和服务访问接口。

4.2.1 接入管理模块

接入管理模块提供跨进程调用功能，为应用层提供核心层功能接入服务。主要包括账户信息查询、账本信息查询、事务操作处理等功能。

（1）账户信息查询

提供内容审核区块链账户的基本信息查询服务。

（2）账本信息查询

提供内容审核区块链的区块、事务详情等查询服务，包括内容审核记录、内容交易、奖惩记录等。

（3）事务操作处理

将用户事务请求提交给区块链网络，包括接口服务能力管理和接口访问权限管理。接口服务能力管理支持接口调用频度设置和事务操作及账本查询缓存设置。接口访问权限管理对不同用户配置不同的访问权限。

4.2.2 节点管理模块

节点管理模块具有节点服务器信息查询、节点服务启动关闭控制、节点服务

配置、节点网络状态监控、节点授权管理等功能。

（1）节点服务器信息查询

提供区块链节点服务器的节点状态信息查询服务。

（2）节点服务启动关闭控制

提供区块链节点服务器的启动和关闭服务。

（3）节点服务配置

提供区块链节点服务器的节点服务能力配置。

（4）节点网络状态监控

提供区块链节点服务器网络连接状态监控。

（5）节点授权管理

提供区块链节点准入准出配置和节点事务处理及账本查询授权配制。

4.2.3 账本应用模块

账本应用模块具有内容发行和交换、逻辑验证、权限控制、合约执行等功能。基于区块链分布式数据存储机制，通过不同节点对账本的共同记录和维护，形成区块链系统中的公共管理、防篡改和可信任机制。区块链应支持持久化存储，多节点拥有完整的数据记录，以及向获得授权者提供真实的数据记录，从而确保相同账本记录的各节点的数据一致性。

账本应用模块具有收集交易数据、生产数据区块、本地数据合法性校验及将校验通过的区块添加到链上等功能。由于数据结构中包含发送人的数字签名，因此可通过已公开的公钥信息，对交易中的数字签名进行验证，实现链上内容的发布和交换、共识前的逻辑验证和共识后的结果验算、多签名权限控制设置、基于合约服务模块执行合约逻辑等功能。

4.2.4 合约服务模块

智能合约是一种可自动执行的数字化协议，可按照预设合约条款自动执行。内容审核区块链可以选择使用智能合约，在网络启动时，部署一套功能强大、结构灵活且支持自定义扩展的智能合约，实现准入控制、身份认证、配置管理、权限管理等功能。原则上由管理员在网络启动时部署全网生效。

系统合约原则上由区块链管理员在网络启动之初部署全网生效。若是在网络运行期间重新部署变更升级，则需要在全网所有节点许可的情况下由区块链管理

员来执行操作。

4.2.5 服务访问接口

服务访问接口为应用层发起的查询、交易、配置和监控等操作提供接口服务。

4.3 核心层

核心层是内容审核区块链系统的核心功能层，包括共识机制、加密算法、数字摘要、数字签名等，为服务层提供基础功能支撑。节点间的共识机制以及基于共识机制的数据和账本记录是区块链系统的基础。加密算法、数字摘要、数字签名等模块保证了区块链系统的安全合规和防篡改能力。

4.3.1 共识机制

共识机制是区块链中各分散节点对事务状态的验证、记录、修改等行为的有效性达成快速共识的基础，为了确保信息的准确性和有效性，区块与区块之间通过共识机制判断数据有效性。共识机制结合容错机制的应用，达成对某一数值或区块链状态的共识。

共识机制具备以下功能：

（1）支持多个节点参与共识和确认。

（2）支持独立节点对区块链网络提交的信息进行有效性验证。

（3）防止任何独立的共识节点未经其他共识节点确认而在区块链系统中进行信息记录或修改。

（4）具有一定的容错性，包括节点物理或网络故障的非恶意错误，节点遭受非法控制的恶意错误以及节点产生不确定行为的不可控错误。

区块链的共识算法主要包括：工作量证明（PoW，ProofofWork）、权益证明（PoS，ProofofStake）、权益授权证明（DPoS，Delegated Proof Of Stake）、实用拜占庭容错（PBFT，Practical Byzantine Fault Tolerance）、权威证明（PoA，Proof of Authority）等。其中，PoW 通过算力竞争获得共识，能源消耗巨大，吞吐量低、延迟过高。PoS、DPoS 需要通过代币数量来控制共识，容易造成代币集中化，使得共识被少数人控制。PBFT 是一种适用于传统分布式系统的拜占庭容错算法，通过三轮广播通信完成共识算法，如图 4-3 所示。

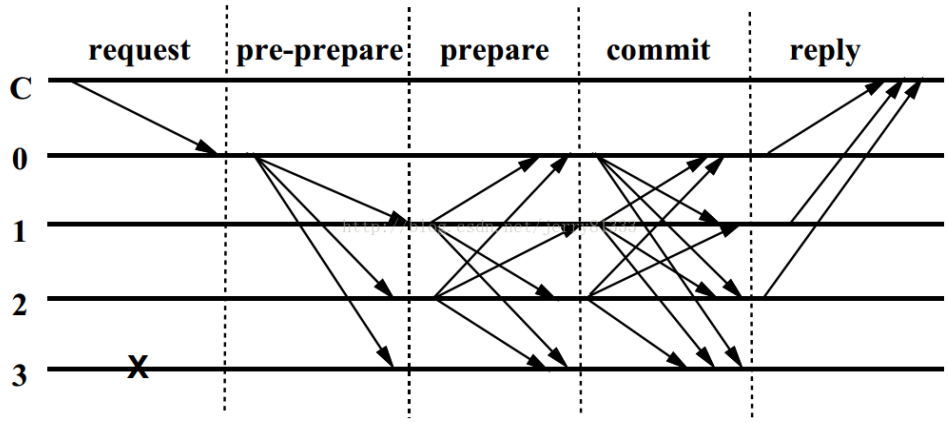

图 4-3PBFT 共识算法

内容审核区块链基于联盟链基础架构，节点通过数字证书技术证明身份且与实体机构一一对应，恶意攻击行为将大大降低，因此，联盟链具有天然的信用基础。同时，联盟链的网络规模和节点数量将受到控制，节点数量会保持在一个稳定的值。和公有链相比，联盟链需要大幅提高可用性，更快地处理网络中的数据，原有低效率高耗能的记账权获取方式将不再适用。因此，内容审核区块链共识算法的需求是：在网络规模相对稳定、参与节点数量可控、存在拜占庭问题的环境下，能够高效处理交易数据，快速完成区块的生成、复制、提交的共识算法。

区块链的共识机制没有完美无缺的，只能根据实际应用需求选择现有算法或者对已有算法进行改进。PBFT 是一种拜占庭容错共识算法，是目前应用于联盟链的相对成熟的共识算法，该算法可以抵御恶意行为的攻击，且具有高度的事务确定性、无分叉快速确认的效果，在有限的节点中共识效率稳定。因此，建议内容审核区块链采用 PBFT 或以 PBFT 为基础进行改进的共识类算法，保证分布式账本的一致性。

但是，PBFT 算法也存在一些不足。首先是算法效率问题，为了保证异步模式的安全性，三阶段广播过程需要消耗较高的通信成本；其次是算法可扩展性问题，由于其三阶段广播过程均需要超过总节点数 2/3 的节点数同意，通信成本会随着节点数量增多而迅速升高，高昂的成本限制了其可扩展性。

为了避免节点数量造成的算法效率骤降问题，内容审核区块链节点应保持在 100 个节点以下。随着内容审核区块链的稳定运行，参与节点的数量有可能不断增加，这种情况下可以通过选择代表或者分组的方式，使参与共识的节点数保持稳定。

4.3.2 加密算法

加密算法是区块链底层安全机制的核心算法，加密算法的安全性和数学难度相关。建议内容审核区块链支持国密算法，例如 SM4、SM7 等对称加密算法和 SM2、SM9 等非对称加密算法。内容审核区块链的链下内容存储建议支持我国商密算法 SM4。

此外，应具备明确的密钥管理方案，确保区块链底层安全机制正常运行。同时，加密算法应具备抵御破解的能力，需定期审核加密算法的安全性，必要时采用更高复杂性的加密算法。

4.3.3 数字摘要

数字摘要用于保证给定的数据明文和摘要不被篡改，对数据的完整性提供保护。该算法通过对任意长度的输入消息进行哈希运算，转化成固定长度的短消息输出，输出值称为哈希值。

内容审核区块链的摘要算法应具备抵御破解的能力，建议支持我国商密算法 SM3，并定期审核摘要算法的安全性。

4.3.4 数字签名

数字签名是指签名者利用私钥对消息进行签名，验签者通过签名者公钥验证签名，实现不可否认的数据源认证。数据签名的功能包括：①对转账交易、合约执行进行签名授权；②通过签名实现权限控制，同时证明授权不可否认；③确保交易数据经过签名后没有也不能被其他任何人修改。

数字签名技术是非对称加密技术和数字摘要技术的结合，包括数字签名和签名验签两个操作。发送方先将待签名的原文通过哈希算法映射成一段哈希值，然后使用私钥对这段哈希值进行签名，验证者利用签名者的公钥对数字签名值和信息原文进行签名验签。

内容审核区块链的签名算法应具备抵御破解的能力，建议支持我国商密算法 SM2 签名算法，并定期审核数字签名所使用的非对称加密算法的安全性，必要时采用更高复杂性的算法。

4.4 基础层

基础层提供了区块链系统正常运行所需要的运行环境和基础组件，包括数据

存储、运行容器、通信网络等，是区块链系统的基础支撑。基础层包括 P2P 网络、分布式存储和计算资源。

4.4.1 P2P 网络

P2P 网络主要为共识达成及数据通信提供底层支持。区块链的网络层本质上是一个 P2P（Peer to Peer，点对点）网络，基于其点对点传输特性，实现分布式网络的联络机制，以及交易和共识信息的接收与广播。网络中的资源和服务分散在区块链节点上，每一个节点作为独立个体进行信息的接收、处理和反馈，这些节点既接收信息也产生信息，节点之间通过维护共同的区块链实现信息的同步。

P2P 网络的安全性和通信效率是保障内容审核区块链整体运作效率的重要因素。内容审核区块链各参与方之间的跨机构通信受现有系统复杂度的影响，例如现有系统为了保障网络安全和内容安全，都会采用内外网多层防火墙隔离的方式，对通信效率有一定的影响。因此，为了保障内容审核区块链各参与方之间的高效通信，建议对现有网络协议进行优化，满足以下要求：①通信时延控制在毫秒级；②具备链路冗余机制保障消息可达；③通讯链路使用 SSL 加密保障消息安全传输，且加密算法可配置。

4.4.2 分布式存储

区块链应支持持久化存储，支持多节点拥有完整的数据记录，支持向获得授权者提供真实的数据记录，确保相同账本记录的各节点的数据一致性。

由于区块链技术的容量限制问题，内容审核区块链分布式存储架构分为链上存储和链下存储两部分，如图 4-4 所示。

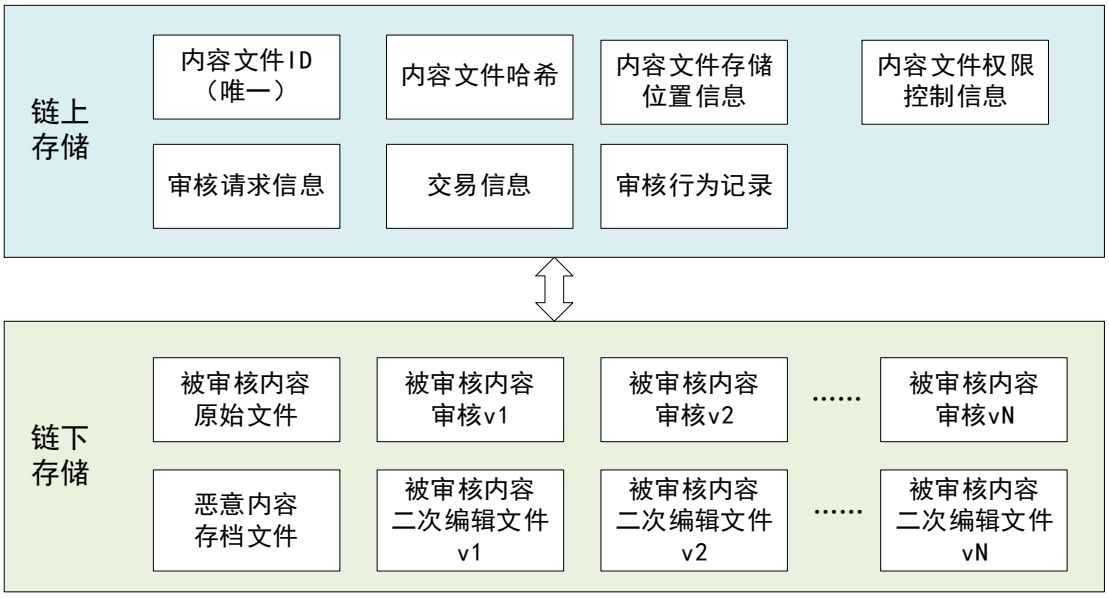

图 1 内容审核分布式存储架构

链上主要存储内容文件 ID、内容文件哈希、内容文件存储位置信息、内容文件权限控制信息、审核请求信息、交易信息、审核行为记录等数据，保障审核、交易、存储、分发等重要信息的完整性和不可篡改性，同时杜绝重要数据的未授权访问。

链下主要存储被审核内容原始文件、审核过程文件、恶意内容存档文件、二次编辑文件等。对于同一视频文件的多次审核应保存全部版本文件以及相应的恶意内容文件（审核问题快照），便于审核记录跟踪溯源。链下存储不建议采用传统的集中式存储，避免单点故障风险。建议采用可寻址的分布式存储系统，例如行星际文件系统（IPFS）、分布式存储系统（Ceph）等。

为了保障内容存储的安全性，避免链下存储的内容受到未经授权的更改，建议对内容文件的完整性进行验证。内容制作机构或内容播出机构在提出内容审核请求时，将内容文件进行哈希计算并将文件哈希上链，内容审核机构获取到媒体文件后也进行哈希计算，并与原始文件哈希进行对比，对比一致可认为内容未被篡改，从而确保内容文件的完整性和一致性。同时，应使用可靠的通信协议确保哈希值不会丢失，在哈希值数据传输时使用可靠的 TCP 协议。

4.4.3 数据结构

本白皮书设计了内容审核区块链系统的主要数据结构，包括视频信息、内容审核请求信息、内容审核结果信息、交易信息和区块信息，为系统开发者提供参

考。具体内容如表 4-4 至表 4-8 所示。

表 4-4 视频信息

属性	类型	描述
vid 视频 ID	字符	视频 ID
hash 哈希	字符	视频文件哈希值
loc 索引	字符	视频文件存储位置索引
pid 提供机构	字符	内容提供机构
tag 标签	字符串	包含人物、事件等内容标签

表 2 内容审核请求信息

属性	类型	描述
vmsg 信息	字符	相关视频信息序列化
hash 哈希	字符	内容审核请求信息哈希值
type 类型	字符串	用于标识请求业务类型
uid 用户 ID	字符	交易用户 ID
expire 过期时间	时间	审核请求有效期
sig 签名	字符串	用户私钥进行信息数字签名

表 4-6 内容审核结果信息

属性	类型	描述
uid 审核人	字符	审核人 ID
hash 哈希值	字符	审核结果信息的哈希值
time 时间戳	时间	审核时间
exp 过期时间	时间	审核有效期
vmsg 信息	字符	原始视频信息序列化
vmsg_new 信息	字符	完成审核的视频信息序列化
t 审核类型	整数	0：针对视频内容审核 1：针对审核机构表现审核
evidence 表现证据	字符串	仅针对审核类型为 1 的审核信息有效。包含涉及的审核机构发布的审核信息，作为审核结果评分依据

属性	类型	描述
Eva 审核结果	整数	对审核类型为 0 的审核信息，0 与 1 代表内容是否通过 对审核类型为 1 的审核信息，其值代表其评分加减
exp 结果解释	字符串	对于审核结果解释说明
cmsg 信息	字符串	因重播重审或其他情况，引用的过去的序列化后的内容审核结果信息
sig 数字签名	字符串	信息数字签名

表 4-7 交易信息数据结构

属性	类型	描述
hash 哈希值	字符	交易的哈希值
msg 信息	字符	审核申请、审核结果等序列化后信息
nonce 随机数	整数	防止重放交易
cid 通道 ID	整数	交易使用的通道信息
sig 数字签名	字符	交易用户使用私钥对交易信息内容进行数字签名

表 4-8 区块信息数据结构

属性	类型	描述
block_height 区块高度	整数	区块链网络中当前区块的高度，也可以理解为第几个区块
hash 哈希值	字符	区块的哈希值
time 时间戳	时间	区块生成时的本地时间
cid 通道 ID	整数	区块使用的通道信息
n_tx 交易量	整数	本区块包含的交易数量
prev_block 父区块	字符	前一个区块的哈希值
mrkl_root 默克尔树根	字符	区块的默克尔树根哈希值
txids 交易 ID	字符串	所有包含的交易哈希值
txs 交易	字符串	所有包含的交易信息

4.5 安全机制

区块链出现至今遭遇了大量的网络攻击，根据区块链安全信息平台 BCSEC 的统计，2011 年到 2019 年上半年，全球范围内因区块链安全事件造成的损失多

达 43.35 亿美元[①]。涉及的安全威胁包括获取服务器权限、修改关键信息、盗取密钥、篡改交易金额、泄露敏感信息等。

内容审核区块链的安全保障采用分层的思想建立逐级防御体系，为区块链各层及层间协议提供保密性、完整性、可用性和隐私保护等安全保障。主要功能包括密钥安全、合约安全、网络安全、存储安全、数据安全等。同时，在保障技术安全的基础上，人员的安全意识培养和安全管理也是区块链安全的重要组成部分。

4.5.1 内容安全

媒体内容安全主要防止媒体内容被非法下载和传播等，是内容审核区块链的关键问题。内容审核区块链可根据内容价值、审核时限、播放范围等需求选择采用以下机制保障内容安全。

（1）访问控制

访问控制是内容安全的基础保护措施，通过配置媒体内容文件的访问策略达到基本的内容安全保护。访问策略包括：允许访问白名单、拒绝访问黑名单、访问次数和访问时间限制、IP 数限制等。

（2）视频加密

视频加密通过对媒体文件加密可有效防止视频内容泄露，避免恶意传播。可根据内容价值、审核时间要求等因素选择使用。

视频加密方法可以使用商业 DRM、HLS 标准加密或其他私有加密算法等。商业 DRM 安全等级较高，能够满足内容提供商的安全要求，但需要支付额外的 License 费用，并需集成特定的 SDK。HLS 标准加密使用 AES-128 对视频内容本身进行加密，同时能支持所有的 HLS 播放器，可免费使用，安全等级较低，但通用性较强。

为保证安全，建议每个媒体文件拥有独立的加密钥匙，这样能有效避免采用单一密钥时，因一个密钥的泄露而引起大范围的安全问题。

提供信封加密机制“密文 Key+明文 Key”，仅“密文 Key”入库，“明文 Key”不落存储，所有过程只在内存中，用完即销毁。

（3）安全下载

安全下载是将媒体文件通过私钥进行二次加密，下载后在播放器 SDK 内部完

[①]BCSEC，区块链安全分析报告，2018。

成视频解密，保障离线视频仅能通过唯一应用进行安全播放。这同样是内容安全防护的可选方式。

4.5.2 数据安全

区块数据是分布在多个节点上的链式结构数据，单一或少部分节点的区块数据篡改不会影响整个区块链的运行。区块链的数据安全风险主要包括：利用数据不可删除的特性进行恶意信息攻击，利用大量垃圾信息进行资源滥用攻击。

针对以上风险，内容审核区块链采用多层次的安全保障机制实现数据安全，包括准入控制、权限控制、数据脱敏、数据加密和隐私保护。

（1）准入控制

用户需要经过审核和身份验证才能加入联盟，通过准入控制机制验证参与者的身份和证书，确认用户对数据的访问权限。

（2）权限控制

结合角色设计和权限控制，实现对数据读写权限的过滤，持有密钥的用户才能解密和访问数据。可对系统资源采用最小权限原则来进行权限管理。

（3）数据脱敏

对于用户资料等敏感数据，采用伪码上链或不上链，只将哈希摘要和少量元数据上链。

（4）数据加密

加密算法保证区块链的安全和不可篡改性。采用算法复杂度较高的对称加密、非对称加密算法，用密码信封完成加密数据分享。

（5）隐私保护

采用承诺系统、零知识证明、同态加密等方式规避隐私暴露。

4.5.3 密钥安全

加密算法是区块链的核心技术之一。在非对称加密算法中，加密时使用公钥，解密时使用私钥。公钥一般是公开的，私钥由个人持有，是用户拥有数字资产的唯一凭证。传统的中心化机构（例如银行等金融机构）可以通过实名认证等手段，实现账户的冻结和恢复。而在区块链网络中，一旦私钥丢失或泄露，可能对区块链系统造成不可估量的影响。因此，私钥安全是区块链安全的重要组成部分。

私钥安全遵从生命周期安全的思想，从私钥的生成安全、存储安全和使用安

全三个环节进行保障。

（1）私钥的生成安全

私钥的质量取决于产生私钥的随机数的质量，高质量的随机数具有不可预测性。随机数分为伪随机数和真随机数两种，伪随机数依靠种子和算法产生，具有可预测性，安全性较差。真随机数基于硬件设计，根据温度、电压、磁场、环境噪声等产生，具有不可预测性。

（2）私钥的存储安全

私钥的存储和使用一般分为软实现和硬实现。软实现使用软件形式进行存储和实现，密钥生成后作为文件或字符串保存在用户终端或托管到服务器，使用时直接或通过口令读取私钥明文到内存。这种方式安全风险较大，容易被复制、窃取或暴力破解，攻击者一旦攻破节点就可以窃取到签名私钥，并使用签名私钥进行签名。

硬实现依托于专用密码安全芯片或密码设备，具有无力保护、敏感数据保护、密钥保护等机制，私钥由专用硬件产生，且不以明文形式出现在密码设备外。密码设备内部存储的密钥具备有效的密钥保护机制，防止探测和非法读取。

（3）私钥的使用安全

私钥在使用一定周期后应该更换密钥。私钥由用户拥有和控制，在风险发生时，可以通过线下实名方式或者权威机构监督的方式找回密钥。

建议采用自建或引入第三方 PKI 基础设施的方式，实现密钥的加密分发、更新、失效等完善的密钥安全管理机制。可以由监管方通过为其他节点注册生成密钥，或采用 CA 证书的管理方式，由监管机构作为 CA 根节点，为其他用户生成子节点 CA 证书，以此标识相关用户身份，达成链上数据合理确权的目的。对于其他用户，也可以为其相对应的下一级子用户生成相关的密钥或 CA 证书，使整个系统实现树状多层级管理结构，由上级节点监管下级节点。对于监管方，通过掌握根密钥或 CA 根证书，实现对链上数据的“穿透式监管”。相关密钥或证书可以根据相关用户涉及的数据安全需要，选用硬件或软件作为载体进行存储。

4.5.4 网络安全

区块链的网络安全主要从 P2P 网络安全和节点安全方面进行提升。

（1）P2P 网络安全

区块链的信息传播采用 P2P 网络，P2P 网络是分布式自组织网络，依赖附近的节点来进行信息传输，需要互相暴露对方的 IP，若网络中存在一个攻击者，就很容易给其他节点带来安全威胁。针对 P2P 网络的攻击有日食攻击、窃听攻击、BGP 劫持攻击、节点客户端漏洞、拒绝服务攻击等。

P2P 网络安全应做到以下几方面：

a.在网络传输过程中，使用可靠的加密算法进行传输，防止恶意攻击者对节点网络进行流量窃取或劫持。如开启 JSONRPC 的节点强制使用 HTTPS 传输，而不是 HTTP 协议。

b.加强对网络中传输数据的有效性、合理性、安全性进行验证，防止出现整数溢出等情况导致出现数据错误。

c.加强节点网络安全性。对重要操作和信息客户端节点做必要的验证。

（2）节点安全

区块链节点的 RPC(Remote Procedure Call,远程过程调用)接口以 JSON-RPC 的形式对外提供调用，节点维护者可以利用节点的 RPC 接口控制节点的行为，如签署和发布交易信息。对于轻节点这类因载体限制无法进行信息全存储或全验证的节点,需通过 RPC 接口进行收发信息或数据请求。由于 RPC 底层为 HTTP 协议，缺乏对调用者的身份验证，RPC 攻击是节点安全的重要风险。

对于必须启用 RPC 端口的节点，应采取修改 RPC 端口号、采用基于 TLS 传输层安全加密协议的 RPC 接口（RPC based on TLS）、设置防火墙策略等方式保障节点安全。

为了保障节点间的通信安全，以及对节点数据访问的安全性，在节点方面做到以下安全控制：

a.节点使用 SSL 连接，保障了通信数据的机密性。

b.引入网络准入机制,可将指定群组的作恶节点从共识节点列表或群组中删除，保障了系统安全性。

c.通过群组白名单机制，保证每个群组仅可接收相应群组的消息，保证群组间通信数据的隔离性。

4.5.5 合约安全

由于区块链具有不可篡改的特点，智能合约一旦发布极难修改，合约的安全

也决定了区块链的安全。

智能合约本质上是代码程序，难免会有考虑缺失导致的漏洞。合约安全威胁主要包括可重入攻击、调用深度攻击、交易顺序依赖攻击、整数溢出攻击等。为保障智能合约的安全，应保证开发安全和全面的安全审计。

（1）安全开发

智能合约安全开发包括开发工具、编译工具及安全编程规范等方面的保障。

a.智能合约应采用规范的工具进行开发和编译，使用安全统一的编译器和链接器选项，以便保证编译器所提供的最新的安全保护机制。同时，应使用同一来源的开发工具集，避免攻击者在软件供应链中植入恶意代码。

b.智能合约的编写应基于严格的安全编程规范。常见的安全编码规范包括弃用一些存在注入或内存破坏危险的第三方函数，采用智能合约的checks-effects-interactions 原则进行合约编写等。

c.在智能合约编码中尽可能少地暴露被外部调用的接口。建议采用进程间隔离的技术将智能合约进程与系统隔离，通过 socket 通信方式远程调用智能合约。

（2）安全审计

在发布智能合约之前，做好全面的安全审计工作，有效防范已知风险。第一，选择对应的虚拟机语言扫描工具对合约代码进行静态扫描，扫描事先建模的漏洞类型；第二，为了避免静态扫描可能产生误报的情况，应邀请安全专家进行交互式代码审计，避免重入攻击、未授权访问攻击、Solidity 开发安全等。

（3）安全测试

智能合约部署前必须经过大量的测试，在模拟环境中对智能合约进行攻击测试，挖掘可被利用的漏洞。

5. 内容审核业务场景

内容审核区块链采用联盟链的方式，将内容提供机构、内容审核机构、内容播出机构和监管机构连接起来，基于区块链共识机制，在机构间实时同步媒体内容的审核请求、审核结果，并更新审核规则，从而实现审核链条在联盟体系内的公开透明，让不同机构间的信息流通阻碍降至最低，从而建立互信、公平、共享的协同审核机制，实现以最终审核效果为导向的可追溯、可定位的高效审核机制。

本白皮书面向广播电视节目和网络视听节目，基于现有媒体内容审核流程及业务特点，制定了基于区块链的内容审核业务场景，包括一般审核、短视频审核、上级二审、定责、纠错和重审六类。本章内容为内容审核系统设计和流程接口设计提供参考。

5.1 一般审核

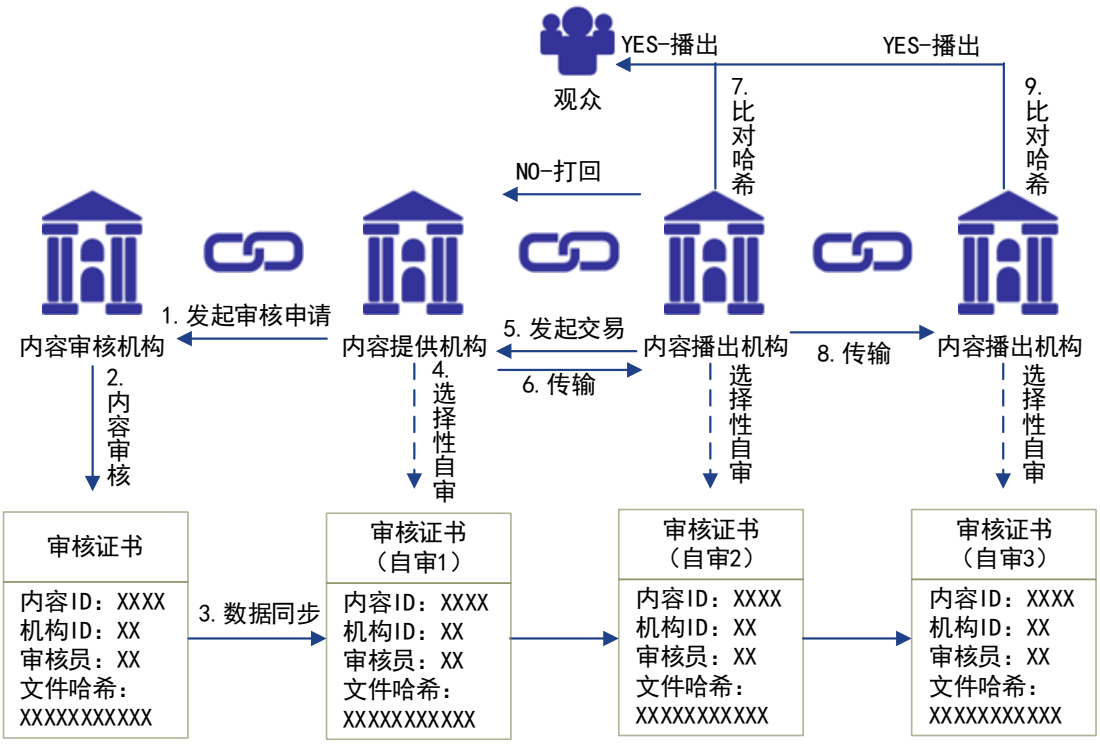

图 5-1 一般审核流程

一般审核是通用的内容审核场景，其流程如图 5-1 所示。由内容提供机构发起内容审核申请，内容审核机构执行内容审核工作。同时，该流程还描述了后续的媒体内容交易流程，由内容播出机构发起内容交易申请，交易完成后，内容播出机构可以选择进行自审或直接取信当前审核结果。

审核证书是内容审核区块链的核心存证，具体字段包括原始内容 ID、完成审核后的内容 ID、内容标签、文件哈希、文件位置信息、审核问题文件 ID、审核问题文件哈希、问题文件位置信息、审核员 ID、审核日期、审核机构签名，详细信息参见 4.1.2 审核服务模块中表 4-3 内容审核结果信息[①]。

一般审核流程详细描述如下：

（1）内容提供机构向内容审核机构发起审核申请，同时将待审核内容进行签名后发送给内容审核机构。

（2）内容审核机构对内容进行线下审核，完成审核后对被审核内容颁发审核证书。

（3）基于联盟链的共识机制，审核证书将同步到联盟内所有机构的节点，一旦上链，该证书信息不可篡改。

（4）内容提供机构在将内容推送/交易给内容播出机构前，可选择对内容进行自审，并将自审的结果同步到链上的审核证书。

（5）内容播出机构发起购买内容交易申请。

（6）内容提供机构将完成审核的内容文件发送给发起内容交易的内容播出机构。

（7）内容播出机构如果采信内容提供机构的审核结果，则将播放文件的哈希与链上审核证书的文件哈希进行对比。若一致，则表示文件无误，可以播出；若不一致，则表示文件有误，将其打回内容提供机构，要求传输正确文件。

（8）内容播出机构如果不信任之前的审核结果，可以选择进行自审或者委托给第三方审核机构进行审核。审核后的结果，无论是否对播放内容有改动，都会同步到联盟内其他机构。

（9）如果下一家内容播出机构选择购买同一内容，重复流程（5）～（8），下一家播出机构同样需要比对哈希确定播放文件无误。

5.2 短视频审核

短视频审核适用于网络短视频内容的审核。按照《网络短视频平台管理规范》的规定，网络短视频平台必须实行节目内容先审后播的制度。网络短视频内容由

[①]后文中描述的短视频审核流程、上级二审流程、定责流程、纠错流程和重审流程中所使用的审核证书与一般审核相同，不再赘述。

个人账户或机构账户上传至内容播出机构（网络短视频平台），内容播出机构按照“先审后播”的制度可以选择自有团队审核或者委托第三方审核机构进行审核。本白皮书面向海量增长的短视频内容审核需求，提出了适用于第三方审核方式的短视频审核流程。短视频审核流程如图 5-2 所示。

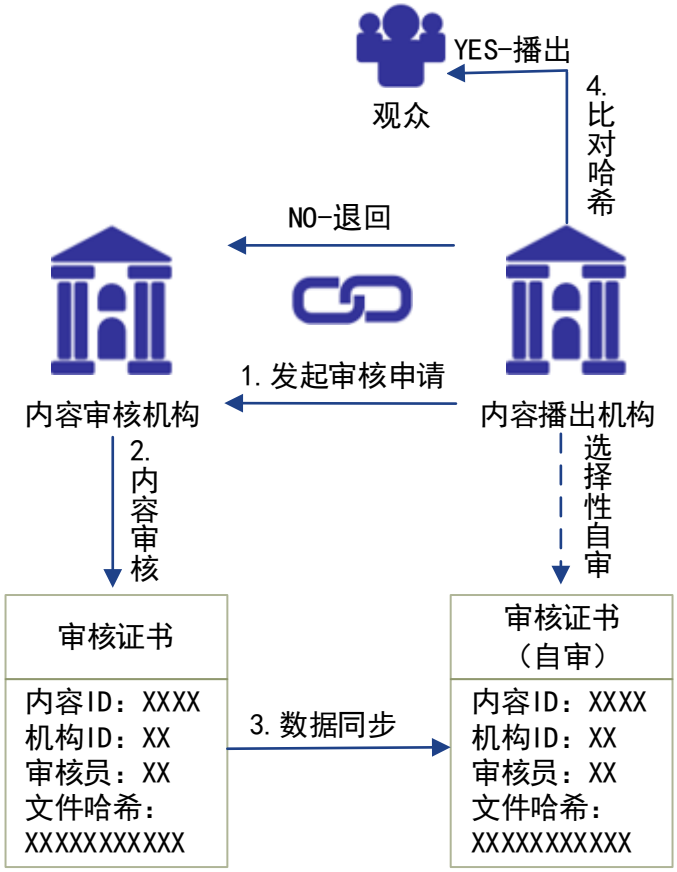

图 5-2 短视频审核流程

短视频审核流程详细描述如下：

（1）内容播出机构向内容审核机构发起审核申请，同时将待审核内容签名后发生给内容审核机构。

（2）内容审核机构对内容进行线下审核，完成审核后对被审核内容颁发审核证书。

（3）基于联盟链的共识机制，审核证书将同步到联盟内所有机构的节点，一旦上链，该证书信息不可篡改。

（4）内容播出机构如果采信内容审核机构的审核结果，则将播放文件的哈希与链上审核证书的文件哈希进行对比。若一致，则表示文件无误，可以播出；若不一致，则表示文件有误，将其打回内容提供机构，要求传输正确文件。

（5）内容播出机构也可以选择对关键内容进行自审以进一步确认审核质量，审核后的结果，无论是否对播放内容有改动，都会同步到区块链。

5.3 上级二审

上级二审适用于《电视剧内容管理规定》规定的内容审核和内容复审，业务流程如图 5-3 所示。图中，内容审核机构（审核委员会）指国务院广播影视行政部门设立的电视剧审查委员会和电视剧复审委员会。

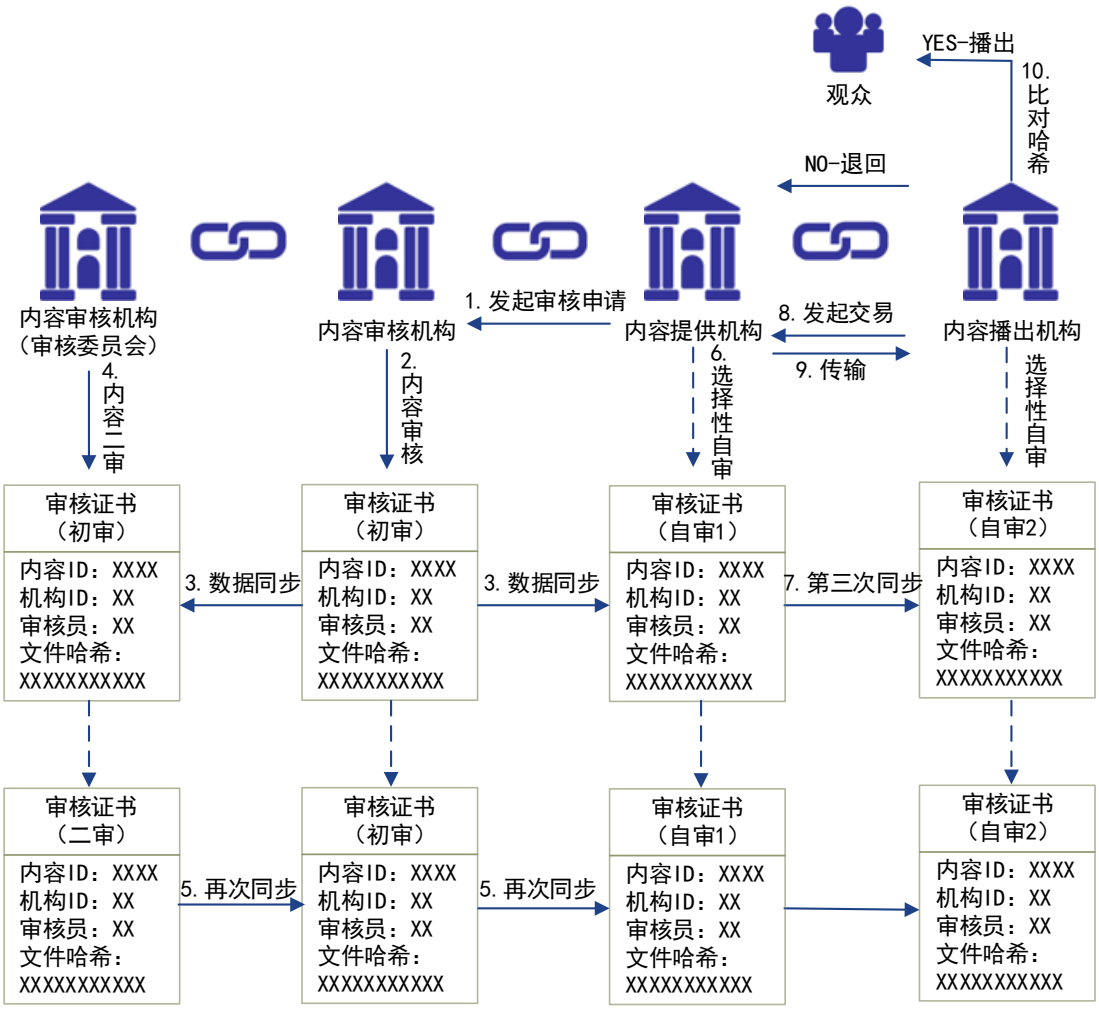

图 5-3 上级二审流程

上级二审的流程描述如下：

（1）内容提供机构向内容审核机构发起审核申请，同时将待审核内容进行签名后发送给内容审核机构。

（2）内容审核机构对内容进行线下审核，完成审核后对被审核内容颁发审核证书。

（3）对于需要国家级内容审核机构（审核委员会）审核的内容，内容审核机构将审核结果同步到区块链上后，由国家级审核机构进行二审，并在链上同步最终审核结果。

（4）基于联盟链的共识机制，审核证书将同步到联盟内所有机构的节点，一旦上链，该证书信息不可篡改。

（5）内容提供机构选择进行自审，将结果同步到区块链，并将内容文件传输给内容播出机构，后续播出流程与一般审核流程相同。

5.4 定责

定责适用于播出过程发现内容不符合审核要求情况下的追责，由监管机构依据审核证书和内容文件进行溯源。定责流程如图 5-4 所示。

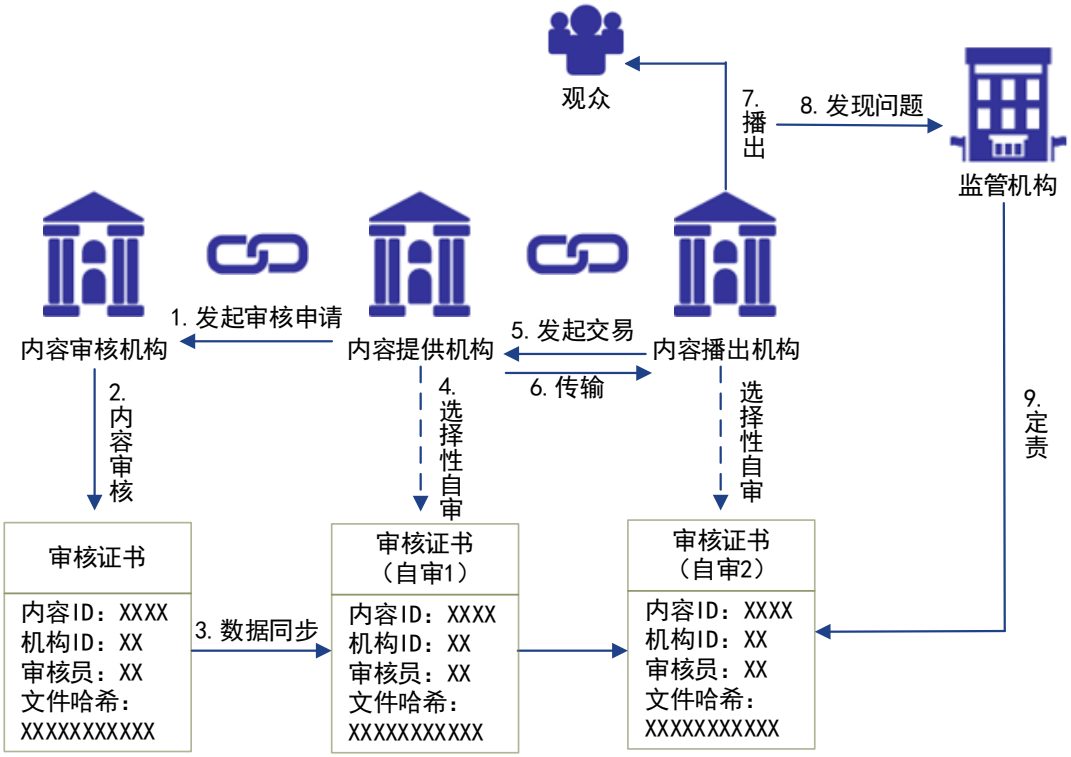

图 5-4 定责流程

定责流程描述如下：

（1）监管机构根据播出事故中所涉及内容 ID、内容标签，检索过往区块中的审核记录。

（2）对于通过上述收集的审核信息，通过审核员 ID、审核机构 ID 或节点地址确定相关负责人。

（3）基于审核信息中的证据，结合应用层的积分评价模块对审核人员及审核机构进行相应的惩罚。

5.5 纠错

当内容提供机构通过自审或第三方审核机构完成内容审核后，通过区块链将自审后审查证书同步给播出机构。此后不管是否播出，其他内容播出机构、内容审核机构或公众发现任何问题，都可将审核失误情况通知内容提供机构或监管机构。本节以内容播出机构审核发现问题为例进行流程说明，如图 5-5 所示。

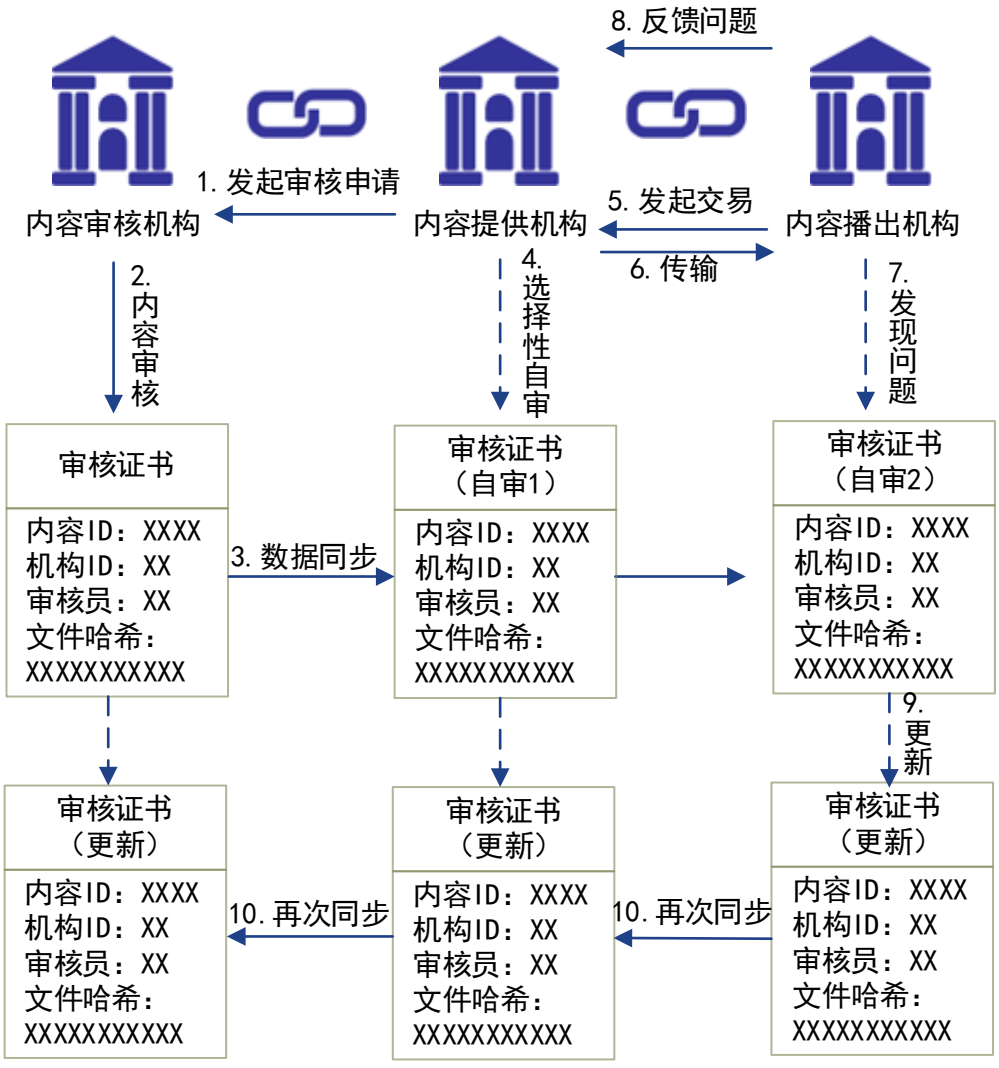

图 5-5 纠错流程

纠错流程说明如下：

（1）内容播出机构通过自审或委托第三方审核机构审核后发现新的问题，将问题反馈给内容提供机构。

（2）内容播出机构采取自审或委托第三方审核机构对内容进行重审，将更新后的审核证书同步到区块链上。

基于区块链数据防篡改、可追溯的特性，针对特定播放文件的所有审核记录都可在链上查证，便于对内容审核机构的专业能力进行客观评价。

5.6 重审

重审适用于内容审核标准（或审核规则）发生变化和内容超过审核有效期的情况。本节以内容审核标准更新为例对重审流程进行描述，重审业务的流程如图5-6所示。

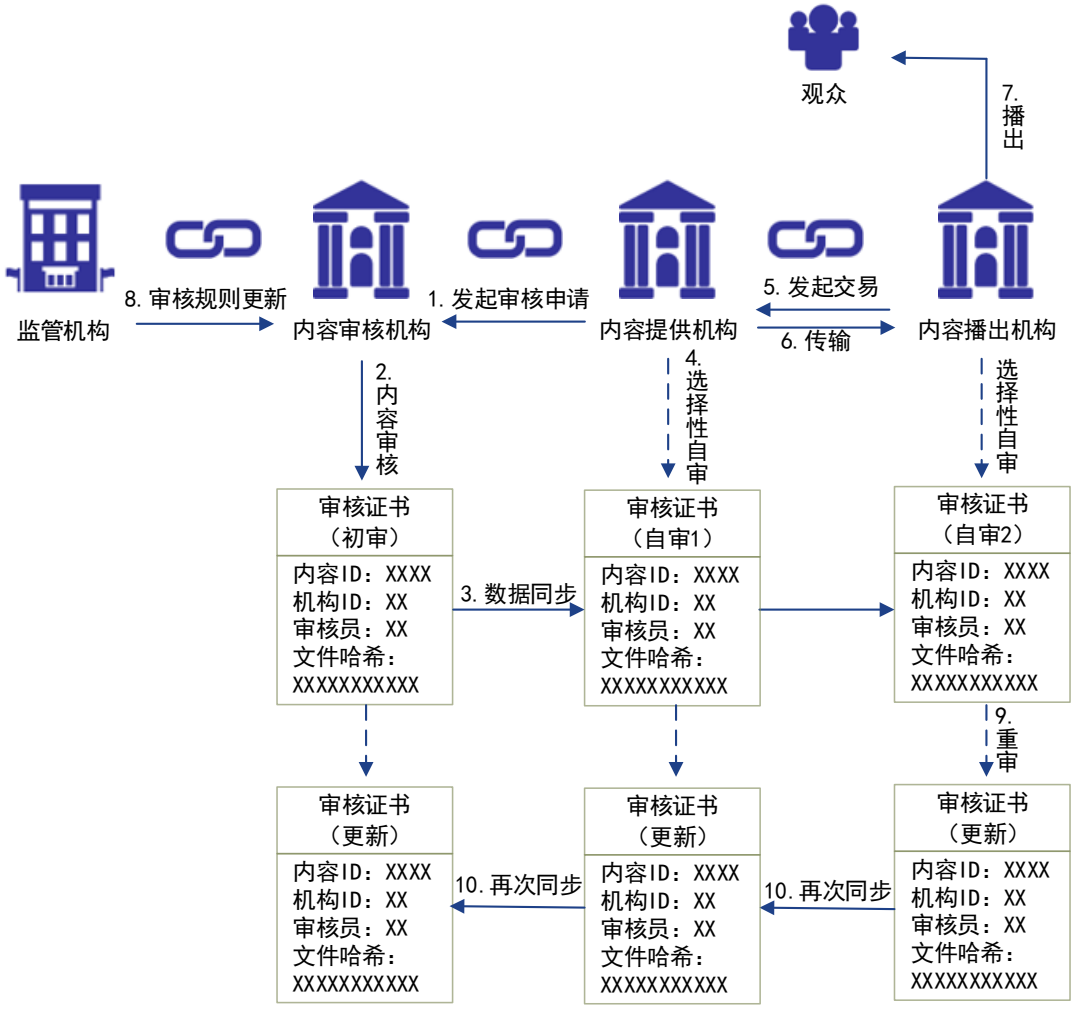

图 5-6 重审流程

重审流程说明如下：

（1）监管机构将更新后的内容审核标准（或审核规则）广播到内容审核区块链。

（2）各内容审核机构收到审核规则变更消息后，基于内容标签进行检索定位，确定需重审的内容。

（3）内容审核机构查询需重审内容相关的播出机构，完成重审业务协商。

（4）内容审核机构根据内容播出机构的重审需求，对相关内容进行重审，完成审核后对被审核内容更新审核证书。

（5）内容提供机构选择进行自审，将结果同步到区块链，后续播出流程与一般审核流程相同。

6. 应用扩展分析

内容审核区块链是为了解决互联网视听内容海量增长造成的内容审核压力，避免重复审核、提升内容审核效率，目前已有相关的应用案例[①]。得益于区块链技术的不可篡改性、分布式共识和可追溯的数据通证，内容审核区块链具备内容安全存储能力、交易行为记录能力、视频完整性验证能力，并在此基础上进行必要的功能扩展，提供高质量的内容加工服务、可信的内容交易服务以及安全的内容分发服务。在逐步提升大数据和人工智能能力的同时，还提供精准的内容聚合和内容传播服务，最终打造成连接内容提供机构、内容审核机构、内容播出机构的良性内容传播服务链。

基于以上分析，本白皮书提出了扩展的媒体内容全流程版权保护和安全共享架构，由应用层和底层区块链构成，如图 6-1 所示。

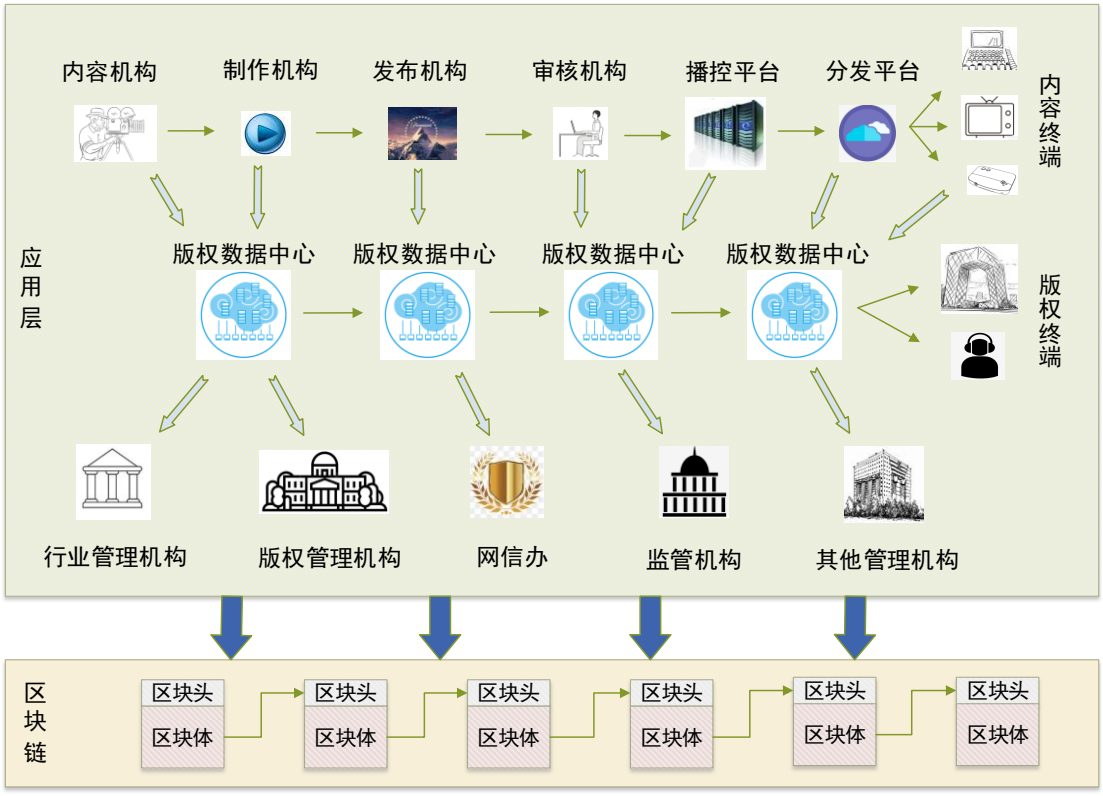

图 6-1 媒体内容全流程版权保护和安全共享架构

应用层面向视频内容生产、内容制作、内容审核、内容播出、内容分发等全流程提供服务，媒体内容经过每个流程都需要将内容文件传输至版权数据中心，

[①]区块链技术应用于内容审核的案例参见《广播电视与网络视听区块链技术应用白皮书（2020）——总体篇》。

并将相关数据同步至区块链，保障数据的真实性与不可篡改性。同时，应用层具备全生命周期的监测监管，向广电管理机构、版权管理机构、网信办、监管机构等提供必要的监管接口，从版权数据中心获取信息数据并进行查验，保障多媒体内容与版权的合法合规。此外，该架构还具备向机构客户和个人用户等提供版权交易功能。

底层区块链主要存储媒体内容相关的交易数据、制作机构数据、播出机构数据、审核机构数据、特征值数据、版权数据等，作为数据通证保障内容服务链的安全可信。媒体内容全流程版权保护和安全共享架构通过进行视频特征值算法扩展，不仅可以保证媒体文件的完整性，还能够保障媒体流的不可篡改性。基于该架构还可以提供除内容审核外的其他服务，包括版权确权、版权交易、监测监管等。

版权确权：版权所有者（机构）将版权内容、版权信息等提交至基于区块链的数字版权保护节点，节点通过认证、审核等流程，确认版权所有者为该内容的版权方，同时加入版权方和认证方不可抵赖的数字签名，并将相关信息传输至版权数据中心，并同步在区块链上。

版权交易：通过版权确权后的内容，发布在版权数据中心，想要获取该内容版权的机构或个人，可以通过区块链查看该内容的历史记录，包括生产时间、版权所有者、版权确权时间等，然后通过版权数据中心进行交易，交易方式包括版权租赁（只有使用权限）、版权购买等。所有交易信息都被记录，并同步在区块链上。

监测监管：通过以区块链技术为基础的多媒体内容特征值提取、对比的方式进行监管，基于内容标签对全网指定内容进行高效控制，实现监管对象的全面化，监管业务的多样化、深入化，监管过程的智慧化、高效化以及结果呈现形式的灵活化。

内容审核区块链将区块链技术和人工智能技术在内容审核领域进行结合应用，有助于加强内容安全保障能力，增强意识形态工作责任制，是保障智慧广电业务良性发展的重要举措。在加快推进落实区块链+AI 赋能的内容审核体系的同时，管理部门要建立以链治链的监管体系，建设多级、层次化的监管链，同步考虑被动式监管和主动式监管的责任，创新监管体制和运行机制，实现内容审核的

网络化、智能化和协同化，实现真正的高效率审核和智慧化监管。

需要注意的是，虽然区块链技术具有一定的内生安全优势，但其面临的安全攻击也层出不穷，涉及应用层的 Web 安全、服务器主机安全、合约层的编码、核心层的节点程序实现及虚拟机的安全设计问题等。因此，在系统开发阶段应基于设计、实现、验证、响应各环节的全生命周期的安全管理，最大程度降低安全风险。

参考文献

[1] 国家广播电视总局职能配置、内设机构和人员编制规定.国家广播电视总局，2018.

[2] 专网及定向传播视听节目服务管理规定.国家新闻出版广电总局令第 6 号，2016.

[3] 关于调整《互联网视听节目服务业务分类目录（试行）》的通告.国家新闻出版广电总局，2017.

[4] 国家广播电视总局关于进一步加强广播电视和网络视听文艺节目管理的通知.广电发〔2018〕60 号，2018.

[5] 国产电视剧片审查审批事项服务指南.国家广播电视总局，2018.

[6] 电视剧内容管理规定.国家广播电影电视总局令第 63 号，2010.

[7] 电视剧拍摄制作备案公示管理办法.国家新闻出版广电总局，2013.

[8] 国家新闻出版广电总局办公厅关于进一步规范网络视听节目传播秩序的通知.新广电办发〔2018〕21 号，2018.

[9] 关于进一步加强网络剧、微电影等网络视听节目管理的通知.广发〔2012〕53 号，2012.

[10] 国家新闻出版广电总局关于进一步完善网络剧、微电影等网络视听节目管理的补充通知.新广电发〔2014〕2 号，2014.

[11] 互联网视听节目服务管理规定.国家广播电影电视总局令第 56 号，2007.

[12] 互联网等信息网络传播视听节目管理办法.国家广播电影电视总局令第 39 号，2004.

[13] 网络视听节目内容审核通则.中国网络视听节目服务协会，2017.

[14] 网络短视频平台管理规范.中国网络视听节目服务协会，2019.

[15] 网络短视频内容审核标准细则.中国网络视听节目服务协会，2019.

[16] 第 45 次中国互联网络发展状况统计报告[R].中国互联网络信息中心(CNNIC)，2020.

[17] 中国区块链技术和应用发展白皮书[R].中国区块链技术和产业发展论坛，2016.

[18] 区块链参考架构（CBD-Forum-001-2017）[S].中国区块链技术和产业发展

论坛，2017.
[19] 区块链数据格式规范（CBD-Forum-002-2017）[S]．中国区块链技术和产业发展论坛，2017.
[20] 区块链隐私保护规范（CBD-Forum-001-2018）[S]．中国区块链技术和产业发展论坛，2018.
[21] 区块链智能合约实施规范（CBD-Forum-002-2018）[S]．中国区块链技术和产业发展论坛，2018.
[22] 区块链存证指南（CBD-Forum-003-2018）[S]．中国区块链技术和产业发展论坛，2018.
[23] 区块链安全生存指南[R]．长亭科技、ConsenSys、比特大陆联合发布，2018.
[24] 区块链安全分析报告[R] .BCSEC，2018.
[25] 区块链电信行业应用白皮书 (1.0 版)[R].可信区块链推进计划，2019.
[26] 区块链白皮书(2019)[R].中国信通院，2019.
[27] 袁勇，倪晓春，曾帅，王飞跃.区块链共识算法的发展现状与展望[J].自动化学报，2018，44(11)：2011-2022.
[28] 高阳阳.基于区块链的内容信息审查研究[J].海峡科技与产业，2019.